Multiresolution Laser Radar Range Profiling with the Expectation-Maximization Algorithm

by

Irene Fung

B.S. Electrical Engineering, Utah State University
(1992)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

May 1994

Signature of Author_____
Department of Electrical Engineering and Computer Science
May 18, 1994

Certified by_____
Professor Jeffrey H. Shapiro
Professor and Associate Head of Electrical Engineering
Thesis Supervisor

Accepted by_____
Professor Frederic R. Morgenthaler
Chairman, Departmental Committee on Graduate Students

Multiresolution Laser Radar Range Profiling with the Expectation-Maximization Algorithm

by

Irene Fung

Submitted to the Department of Electrical Engineering and Computer Science
on May 18, 1994, in partial fulfillment of the requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

In pulsed-imager operation of a coherent laser radar system, a range image is generated by the time-of-flight measurements between transmitted and reflected pulses from raster-scanning a target region. Previous research work in planar-range profiling has shown the expectation-maximization (EM) algorithm to be a computationally-simple procedure for finding the maximum-likelihood (ML) estimates of the azimuth and elevation range slopes and the range intercept. This thesis continues the development by extending the planar-case fitting to a generalized parametric profiling problem. In particular, a multiresolution wavelet basis is introduced to the ML fitting of the range profile at a sequence of increasingly fine resolutions. The weights associated with the EM algorithm are used to determine the stopping point of progressively coarse-to-fine estimation. Fine-scale variations in the estimated profile can be suppressed, resulting in a minimal-dimensionality, wavelet-fit profile. The performance of this estimation scheme is evaluated via computer simulations.

Thesis Supervisor: Professor Jeffrey H. Shapiro
Title: Professor and Associate Head of Electrical Engineering

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to all the people who had given me huge amounts of support for the research and preparation of this thesis.

I am deeply appreciative to my parents, Raymond and Virginia Fung, for all they have done for me during my life. Their sacrifices provided me the opportunity of pursuing a wonderful education at MIT and their love and support has helped me endure all the difficulties I faced in my studies for this degree. I want to thank my two sisters, Shirley and Elaine Fung, for their friendship and confidence in me, and all the moments of sharing our joys and sorrows.

I am deeply grateful to my academic and research advisor, Jeffrey Shapiro, for all his advice and guidance during my career at MIT. His deep knowledge in laser radar and estimation theory has helped me push through a lot of technical stumbling blocks throughout the research. His clarity of thought, diligence in working and uncompromising commitment to excellence were very good examples for me. I am especially thankful for his sensitivity, patience, accessibility and advice during the last stage of the production of this thesis.

I want to express my sincere gratitude to Gideon Lee, for all the understanding, patience, encouragements and sacrifices he has given me. His valuable ideas and advice on the implementation of the simulation program has helped me to improve its performance tremendously. Thanks for standing by my side all this time.

I would like to thank all the good friends here at MIT. I am grateful for the companionship, support and understanding of Ada Yau, Kai Pak Chan, Kinhong Kan during these few months when we were all working on our theses. To Jocelyn Chen, Phoebe Fu, Kingchung Yu, Katherine Tso, Sinyan Law, Bernard Wong and Jenny Siu — thanks for letting me use your Cray times for my simulations.

I am appreciative to everyone in the Automatic Target Recognition Group at MIT for their technical resources and advice. The presentations and discussions have been very informative and rewarding.

Finally, I want to express my appreciation to the U.S. Government for the support I have received through the U.S. Air Force Office of Scientific Research Grant F49620-93-1-0604. which has made this research and the completion of my graduate studies possible.

Contents

List of Figures

Chapter 1

Introduction

Coherent laser radars are capable of collecting intensity, range, or Doppler images by raster scanning a field of view [1, 2]. In pulsed-imager mode, the range image is produced by measuring the time-of-flight between the transmitted pulse for each pixel and the peak-intensity of the video-detected return waveform. These range images are often degraded by the combined effects of laser speckle and local-oscillator shot noise. The fact that most reflecting surfaces are rough, when measured on the scale of a laser wavelength, causes constructive and destructive interference in the reflected light known as speckle [3]. On the other hand, shot noise is introduced in the optical detection process which imposes a finite carrier–to–noise ratio (CNR) on the signal [4].

Interest in the statistics of peak-detecting coherent laser raders has stimulated research in single-pixel statistics [5], target-detection for 2-D imagers [7-9] and 3-D imagers [10-12]. In a recently completed doctoral thesis by T.J. Green, maximum-likelihood (ML) estimation is used to fit a planar range profile to a frame of laser radar range data. In particular, the expectation-maximization (EM) algorithm is employed to obtain the ML estimate. Simulation results have confirmed this technique to be computationally simple with good noise suppression. However, the presumption of a planar-range profile is unduly restrictive.

This thesis will extend the planar range profiling work to a more general scheme. In particular, a multiresolution wavelet basis is incorporated into this parametric

range profiling approach. Starting from the single-pixel statistical model for range measurements from a peak-detecting, pulsed-imager laser radar [5], the EM algorithm is again employed to achieve the ML estimate. The general theory of this approach and results on simulated range data will be presented.

This thesis will also address two of the major issues that arise from this multiresolution range profiling process. Specifically, the termination criterion for the multiresolution process and the fine-tuning of the estimated profiles are presented.

The remainder of this thesis is organized as follows. Chapter 2 describes the single-pixel statistical model used for the derivation of the multiresolution range profiler. Chapter 3 discusses the range profiling estimation problem and the planar range profiler. Chapter 4 describes the general theory of the multiresolution range profiler and studies the performance of this estimation process through simulation. Chapter 5 provides a method of terminating the multiresolution process. Chapter 6 introduces the piecewise-smoothing method for fine-tuning the estimated profile. It also leads to the discussion of using maximum a posteriori probability (MAP) estimation as an alternate smoothing scheme. In Chapter 7, the major conclusions of this work are summarized.

Chapter 2

Background

In this chapter, the single-pixel statistical model for the range measurements of a pulsed-imager laser radar is presented. This model serves as the basis for the derivation of the range profiling theory to be addressed subsequently.

2.1 Single-Pixel Statistics

In the pulsed-imager mode, a laser pulse is transmitted for each pixel across the raster. The reflected light collected then undergoes optical heterodyne detection, intermediate-frequency filtering, video detection, and peak detection, as shown in Fig. 2-1. Each range pixel represents the time delay between the peak of the transmitted pulse and the peak intensity of the video-detected return waveform from an optical heterodyne receiver. These range images are subject to fluctuations arising from the combined effects of laser speckle and local-oscillator shot noise. The former is due to the rough-surfaced nature of most reflecting surfaces measured on the scale of a laser wavelength [3], and the latter is due to the fundamental noise encountered in optical heterodyne detection [4].

Speckle degrades laser radar range images by creating range anomalies. An anomaly occurs when a deep target-return fade and a strong noise peak conspire to give a range measurement very far from the true range value [5]. Fig. 2-2 shows the difference between anomalous and non-anomalous behavior.

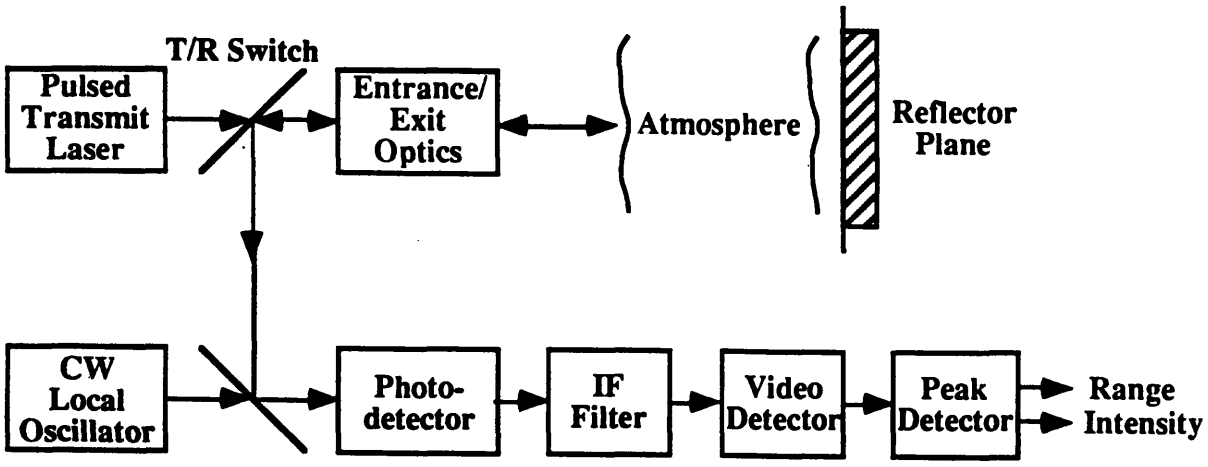

Figure 2-1: Block diagram of a monostatic, shared-optics coherent laser radar.

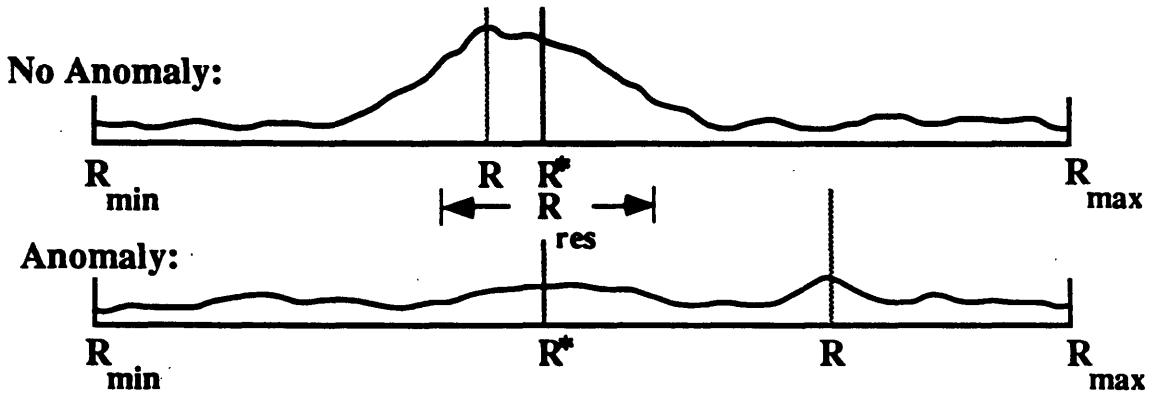

Figure 2-2: Peak detection range measurement examples showing nonanomalous and anomalous behavior. Herein, R^* is the true range value, R is the measured range, R_{res} is the range resolution, and $\Delta R \equiv R_{max} - R_{min}$ is the range uncertainty interval.

Let r denote the range value of a given pixel whose true range value is r^*. The conditional probability density function (PDF) for $r = R$, given $r^* = R^*$, is [11]

$$p_{r|r^*}(R \mid R^*) = [1 - \Pr(A)] \frac{\exp\left(-\frac{(R-R^*)^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}} + \frac{\Pr(A)}{\Delta R},$$

$$\text{for} \quad R_{min} \leq R, R^* \leq R_{max}. \tag{2.1}$$

Here, $\Pr(A)$ is the anomaly probability and δR is the local range accuracy. ΔR is the width of the laser radar's range uncertainty interval, $\mathcal{R} = [R_{min}, R_{max}]$, which is assumed to be much larger than δR. The first term on the right of Eq. 2.1 represents the local range behavior. It is the probability of a nonanomalous measurement times a Gaussian probability density with mean equal to the true range, R^*, and standard deviation equal to the local range accuracy, δR. The second term represents the global range behavior. It is the anomaly probability times a uniform probability density over the entire range-uncertainty interval.

Let us assume R_{res} to be the radar's range resolution — roughly $cT/2$ for a pulse of duration Ts, where c is the speed of light [5]. Then in terms of the radar's carrier-to-noise ratio,

$$\text{CNR} \equiv \frac{\text{average radar return power}}{\text{average local-oscillator shot noise power}}, \tag{2.2}$$

the range resolution R_{res}, and the number of range resolution bins $N \equiv \Delta R / R_{res}$, the local range accuracy and the probability of anomaly can be written as

$$\delta R \approx \frac{R_{res}}{\sqrt{\text{CNR}}}, \tag{2.3}$$

and

$$\Pr(A) \approx \frac{1}{\text{CNR}} \left(\ln(N) - \frac{1}{N} + 0.577\right). \tag{2.4}$$

These are valid for the interesting regime of $N \gg 1$ and $\text{CNR} \gg 10$. More exact

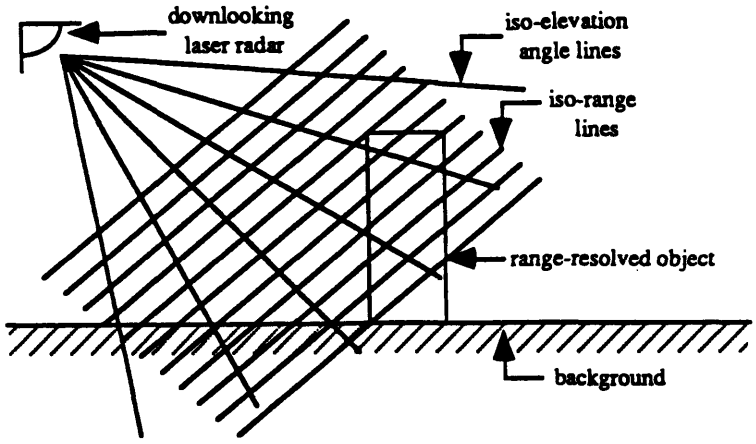

Figure 2-3: Downlooking geometry for a laser radar measurement

results can be found in [5].

For the typical case of monostatic operation, the CNR can be obtained from the resolved speckle-reflector radar equation [6],

$$\mathrm{CNR} = \frac{\eta P_T}{h\nu B}\frac{\rho A_R}{\pi R^{*2}}\varepsilon \exp(-2\alpha R^*), \tag{2.5}$$

where η is the radar receiver's photodetector quantum efficiency; P_T is the peak power of the radar transmitter; $h\nu$ is the photon energy at the radar's laser wavelength, $\lambda = c/\nu$; B is the radar receiver's IF bandwidth; ρ is the reflectivity of the pixel under consideration; A_R is the area of the radar's entrance-pupil; ε is the product of the radar's optical and heterodyne efficiencies; and α is the atmospheric extinction coefficient, assumed to be constant along the propagation path.

2.2 Scene Geometry

The processors used in background range-plane estimation and fine-range target-detection work [11] have assumed a downlooking geometry, i.e., the laser radar is observing a ground-based scene from above, as shown in Fig. 2-3. This scene geometry will be used for the multiresolution range profiler. Furthermore, there will be no restriction on the background-profile and the existence of a spatially resolved target.

The measured data is a range image, $\mathbf{r} = \{r_{jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$, obtained from a single raster scan. This $\mathbf{r}$ is the observation vector in our case since we are interested in range-only measurements[1]. The pixel spacing is assumed to be sufficiently large so that each pixel measurement is statistically independent given the true range profile. Hence the joint PDF of the range data is just the product of the single pixel conditional PDFs as discussed in the next section. Our multiresolution range profiler will be developed based on this statistical model.

[1]We will not consider the associated intensity image generated by the laser radar. Interested readers may refer to [11] for work on joint range-intensity processing.

Chapter 3

Range Profile Estimation

In this section, we describe the profile estimation problem and present a general framework of the maximum-likelihood (ML) range-profile estimation with the expectation-maximization (EM) algorithm. The planar-range profiling technique is introduced as an example and extended to a generalized parametric profiling problem.

3.1 Range Profile Estimation Problem

Suppose we use a uniformly spaced raster scan to collect a $J \times K$-pixel range image of some field of view, $\{r_{jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$. We shall assume that the pixel spacing is large enough to ensure uncorrelated speckle on each radar return, so that the $\{r_{jk}\}$ are statistically independent, given the $\{r^*_{jk}\}$. It is then convenient to assemble the range data into a JK-D column vector,

$$\mathbf{r} \equiv \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_J \end{bmatrix}, \qquad \text{where} \quad \mathbf{r}_j \equiv \begin{bmatrix} r_{j1} \\ r_{j2} \\ \vdots \\ r_{jK} \end{bmatrix}, \qquad \text{for} \quad 1 \leq j \leq J. \tag{3.1}$$

Thus the joint probability density for $\mathbf{r} = \mathbf{R}$ to occur is given by

$$p_{\mathbf{r}|\mathbf{r}^*}(\mathbf{R} \mid \mathbf{R}^*) = \prod_{j=1}^{J} \prod_{k=1}^{K} \left\{ [1 - \Pr(A)] \frac{\exp\left(-\frac{(R_{jk} - R^*_{jk})^2}{2\delta R^2} \right)}{\sqrt{2\pi\delta R^2}} + \frac{\Pr(A)}{\Delta R} \right\} \tag{3.2}$$

The laser-radar range profiling problem is to find the "optimal" estimate of the true-range vector $\mathbf{r}^*$ given the range-data vector $\mathbf{r}$. The maximum-likelihood (ML) range estimator appears to be a reasonable approach,

$$\hat{\mathbf{r}}^*_{ML}(\mathbf{R}) \equiv \arg \max_{\mathbf{R}^*}(p_{\mathbf{r}|\mathbf{r}^*}(\mathbf{R}|\mathbf{R}^*)). \tag{3.3}$$

One advantage of using the ML estimator is that no prior knowledge of the statistical behavior of the true-range vector is necessary. It estimates the true-range vector as the $\mathbf{R}^*$ that maximizes the likelihood of getting the range-data vector $\mathbf{R}$. However, this also implies that the best ML estimate that can be obtained is the raw data itself!

$$\hat{\mathbf{r}}^*_{ML}(\mathbf{R}) = \mathbf{R}. \tag{3.4}$$

So ML estimation without a "regulant" condition does not suppress anomalies at all. This can be disastrous because the range anomalies can occur on more than 10% of the pixels even at reasonable CNR. We will show later how the EM algorithm can solve this anomaly problem.

3.2 Planar Range Profiling

In the recent planar range-profiling work [12], the background is modeled as an unknown range plane whose true range values are,

$$r^*_{jk} = x_1 j + x_2 k + x_3, \qquad \text{for} \quad 1 \leq j \leq J, 1 \leq k \leq K. \tag{3.5}$$

Defining a parameter vector,

$$\mathbf{x} \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{3.6}$$

and using the implied PDF,

$$p_{\mathbf{r}|\mathbf{x}}(\mathbf{R}|\mathbf{X}) = p_{\mathbf{r}|\mathbf{r}^*(\mathbf{x})}(\mathbf{R}|\mathbf{R}^*(\mathbf{X})) \tag{3.7}$$

we have

$$\hat{\mathbf{x}}^*_{ML}(\mathbf{R}) \equiv \arg\max_{\mathbf{X}}(p_{\mathbf{r}|\mathbf{x}}(\mathbf{R}|\mathbf{X})). \tag{3.8}$$

Here, $\mathbf{x}$ is treated as an unknown, non-random parameter vector which is to be estimated give the range image $\mathbf{r}$. To achieve that, we employ maximum-likelihood estimation via the expectation-maximization (EM) algorithm.

The EM algorithm is used to develop an iterative procedure that produces a sequence of tentative estimates with monotonically increasing likelihoods. With a sufficiently good initial estimate, the EM algorithm will climb the highest hill of the likelihood surface, yielding the desired ML estimate. Furthermore, it has been shown that the performance of this procedure is satisfactory with anomaly probability as high as 0.5.

This algorithm works well because if the background profile is approximately planar, we are only fitting three parameters with a large number of pixels. The fact that 10% of the total pixels are anomalous does not have a significant effect on degrading the performance of the estimate. The same idea of anomaly-suppression in a multiresolution range profiler will be discussed in the next section.

3.3 Parametric Range Profiling

There is no reason to confine the ML profiling approach described above to planar cases only. In this section, a more general parametric ML profiler with the EM algorithm will be discussed.

Suppose we define $\{\boldsymbol{\phi}_q : 1 \leq q \leq Q \equiv JK\}$ to be an arbitrary orthonormal column-vector basis for Q-D vector space, and we let

$$\mathbf{H} \equiv [\ \boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \cdots \quad \boldsymbol{\phi}_Q\ ] \tag{3.9}$$

be the $Q \times Q$ coordinate-transformation matrix generated by this basis. Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_Q \end{bmatrix} \tag{3.10}$$

be the Q-D parameter vector so that the true-range vector can be written as

$$\mathbf{x} \equiv \mathbf{H}^T \mathbf{r}^*. \tag{3.11}$$

Since $\{\boldsymbol{\phi}_q\}$ forms an orthonormal basis, $\mathbf{H}$ is invertible and $\mathbf{H}^{-1} = \mathbf{H}^T$ is its inverse. Thus estimating $\mathbf{r}^*$ from $\mathbf{R}$ is equivalent to estimating $\mathbf{x}$ from $\mathbf{R}$, and we have

$$\hat{\mathbf{r}}^*_{ML}(\mathbf{R}) = \mathbf{H}\hat{\mathbf{x}}_{ML}(\mathbf{R}) \tag{3.12}$$

and

$$\hat{\mathbf{x}}_{ML}(\mathbf{R}) = \mathbf{H}^T \hat{\mathbf{r}}^*_{ML}(\mathbf{R}). \tag{3.13}$$

Suppose, for a particular $\mathbf{H}$, we know that only the first P dimensions of $\mathbf{x}$ are nonzero, i.e., the true range profile can be characterized by a P-D vector,

$$\mathbf{x}_P \equiv \mathbf{H}_P^T \mathbf{r}^*, \tag{3.14}$$

where

$$\mathbf{H}_P \equiv [\ \boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \cdots \quad \boldsymbol{\phi}_P\ ]. \tag{3.15}$$

It will be convenient to introduce

$$\mathbf{H}_P^c \equiv [\ \boldsymbol{\phi}_{P+1} \quad \boldsymbol{\phi}_{P+2} \quad \cdots \quad \boldsymbol{\phi}_Q\ ] \tag{3.16}$$

so that

$$\mathbf{H} \equiv [\ \mathbf{H}_P \quad \mathbf{H}_P^c\ ], \tag{3.17}$$

and

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_P \\ \mathbf{x}_P^c \end{bmatrix} = \begin{bmatrix} \mathbf{H}_P^T \mathbf{r}^* \\ \mathbf{H}_P^{cT} \mathbf{r}^* \end{bmatrix}, \tag{3.18}$$

where

$$\mathbf{x}_P^c = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3.19}$$

As in the planar case, we are now set up for the ML range profiler. The joint conditional probability density for $\mathbf{r} = \mathbf{R}$ given that $\mathbf{x}_P = \mathbf{X}_P$ is then

$$p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P) = \prod_{q=1}^{Q} \left\{ [1 - \Pr(A)] \frac{\exp\left(-\dfrac{(R_q - (\mathbf{H}_P \mathbf{X}_P)_q)^2}{2\delta R^2} \right)}{\sqrt{2\pi\delta R^2}} + \frac{\Pr(A)}{\Delta R} \right\}, \tag{3.20}$$

where R_q is the qth component of $\mathbf{R}$ from Eq. 3.1.

From the results of the planar profiling problem [12], we can expect to achieve a substantial amount of anomaly suppression by choosing $P \ll Q$. The fact that only 10% of 1000 pixels are anomalous will not affect the estimation result if a small number of parameters is used. The disadvantage to this is the loss of resolution and

detail which may be essential for target recognition. The assumption of a planar background profile is removed in this generalized range profiler, meaning that both the background and the target can be estimated at the same time.

3.4 Maximum-Likelihood Estimation

The maximum-likelihood estimate $\hat{\mathbf{x}}_{PML}$, by definition, is the $\mathbf{X}_P$ that maximizes $p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P)$, given a particular range-data vector $\mathbf{R}$. Since probability densities are always non-negative, it is often easier to maximize the logarithm of the likelihood function. Thus the ML estimate satisfies

$$\frac{\partial}{\partial X_q} \ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P)]\Big|_{\mathbf{X}_P=\hat{\mathbf{x}}_{PML}} = 0, \qquad 1 \leq q \leq P. \tag{3.21}$$

This is the necessary condition for an extremum to occur at $\mathbf{X}_P = \hat{\mathbf{x}}_{PML}$. Performing the indicated differentiation on the density from Eq. 3.20 leads to the same necessary condition as that of nonlinear weighted least-squares estimation, namely,

$$\mathbf{H}_P^T \mathbf{W}_P(\mathbf{X}_P)(\mathbf{R} - \mathbf{H}_P\mathbf{X}_P)|_{X_P=\hat{\mathbf{x}}_{PML}} = \mathbf{0}, \tag{3.22}$$

where the $P \times P$ weight matrix is given by

$$\mathbf{W}_P(\mathbf{X}_P) \equiv \text{diag}[w_q(\mathbf{X}_P)]. \tag{3.23}$$

Here, for $1 \leq q \leq P$, the qth weight is

$$w_q(\mathbf{X}_P) = \frac{[1 - \Pr(A)]\dfrac{\exp\left(-\dfrac{[R_q - (\mathbf{H}_P\mathbf{X}_P)_q]^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}}{[1 - \Pr(A)]\dfrac{\exp\left(-\dfrac{[R_q - (\mathbf{H}_P\mathbf{X}_P)_q]^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}} + \dfrac{\Pr(A)}{\Delta R}}. \tag{3.24}$$

Note that all the weights are proper fractions, $0 < w_q \leq 1$ if $\Pr(A) < 1$. In fact, each weight is the conditional probability that the associated pixel is not anomalous given

that the parameter vector is $\mathbf{X}_P$. Moreover, if the anomaly probability is very small so that $w_q \approx 1$ for all q, then Eq. 3.22 becomes linear and its solution is easily shown to be

$$\hat{\mathbf{x}}_{PML} = (\mathbf{H}_P^T \mathbf{H}_P)^{-1} \mathbf{H}_P^T \mathbf{R}, \tag{3.25}$$

which, due to the orthogonal nature of the $\mathbf{H}$ basis, can be further reduced to become

$$\hat{\mathbf{x}}_{PML} = \mathbf{H}_P^T \mathbf{R}. \tag{3.26}$$

Unfortunately, laser radars often operate in a regime where the number of anomalies is substantial and thus the nonlinear nature of the range estimation problem cannot be ignored. The expectation-maximization (EM) algorithm provides an iterative approach to solving this nonlinear problem and will be discussed in the following section.

3.5 Expectation-Maximization Algorithm

The EM algorithm is very well-suited to ML estimation problems whose observation vectors constitute incomplete data [16]. This includes situations where there is some unobserved data vector of which only a part is the observation vector available for processing. In the context of our ML range profiling, the natural complete data vector is

$$\mathbf{y} = \begin{bmatrix} \mathbf{r} \\ \mathbf{a} \end{bmatrix}, \tag{3.27}$$

where $\mathbf{r}$ is the range-measurement vector — our observations — and

$$\mathbf{a} \equiv \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_J \end{bmatrix}, \quad \text{with} \quad \mathbf{a}_j \equiv \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jK} \end{bmatrix}, \quad \text{for} \quad 1 \leq j \leq J, \tag{3.28}$$

is the anomaly data — the missing part of the complete data. Here,

$$a_{jk} = \begin{cases} 0 & \text{if } r_{jk} \text{ is anomalous,} \\ 1 & \text{if } r_{jk} \text{ is not anomalous.} \end{cases} \tag{3.29}$$

If the complete data $\mathbf{y}$ were available for estimating $\mathbf{x}$, we could identify and suppress the anomalous pixels and reduce the ML estimation problem to a linear task involving only the nonanomalous pixels. Since $\mathbf{a}$ is not directly observed, this convenient way of eliminating the anomalous data is impossible. Instead of having the anomaly data vector, the ML estimate of $\mathbf{x}$ based on $\mathbf{r}$ deals with the possibility of anomalous pixels in a statistical manner, inevitably making it a nonlinear task. However, the linearity of the complete data estimation task makes the EM algorithm computationally simple.

Starting from any initial estimate of the parameter $\hat{\mathbf{x}}_P(0)$, the EM algorithm for solving Eq. 3.22 will produce a sequence of estimates, $\{\hat{\mathbf{x}}_P(n) : n = 0, 1, 2, \ldots\}$ through a series of iterative expectation and maximization steps. The associated likelihood sequence, $\{p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R}|\hat{\mathbf{x}}_P(n)) : n = 0, 1, 2, \ldots\}$, is monotonically increasing. Hence the EM algorithm will climb a hill on the surface $p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P) : \mathbf{X}_P \in \mathcal{X}$, where $\mathcal{X}$ is the set of possible parameter vectors.

A good initial estimate will place the EM algorithm somewhere on the highest hill so that the global maixmum can be achieved. A linear least-squares initial estimate should be sufficient for scenarios with low anomaly probabilities, $\Pr(A) \leq 0.1$. Unfortunately, for most general cases, such initialization is often not reliable to locate the global maximum. The recursive EM algorithm is suggested as an alternative initialization process [12] as discussed later in Sec. 4.4.

Initialization

An arbitrary initialization can be used in the EM algorithm, though whether or not it converges to the correct likelihood maximum depends on the quality of the initial estimate [12]. To be explicit, we assume the least-squares initialization,

$$\hat{\mathbf{x}}_P(0) \equiv (\mathbf{H}_P^T \mathbf{W}_P(0) \mathbf{H}_P)^{-1} \mathbf{H}_P^T \mathbf{W}_P(0) \mathbf{R}, \tag{3.30}$$

where $\mathbf{W}_P(0)$ is the initial $P \times P$ matrix,

$$\mathbf{W}_P(0) \equiv \text{diag}[w_q(0)], \qquad \text{with } w_q(0) = 1, \quad \text{for } 1 \leq q \leq P. \tag{3.31}$$

Here, we have assumed that none of the range data is anomalous and the solution is the same as the one given in Eq. 3.26.

Update Procedure

After the nth step of the EM algorithm, for $n = 1, 2, 3, \ldots$, we have available the current estimate, $\hat{\mathbf{x}}_P(n)$, and its associated weight matrix, $\mathbf{W}_P(n) = \text{diag}[w_q(n)]$. The EM algorithm then updates this estimate to $\hat{\mathbf{x}}_P(n+1)$ by the following two steps:

The *expectation* step updates the weights according to

$$w_q(n+1) = \frac{[1 - \Pr(A)] \dfrac{\exp\left(-\dfrac{[R_q - (\mathbf{H}_P \hat{\mathbf{x}}_P(n))_q]^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}}}{[1 - \Pr(A)] \dfrac{\exp\left(-\dfrac{[R_q - (\mathbf{H}_P \hat{\mathbf{x}}_P(n))_q]^2}{2\delta R^2}\right)}{\sqrt{2\pi\delta R^2}} + \dfrac{\Pr(A)}{\Delta R}},$$
$$\text{for} \quad n = 0, 1, 2, \ldots, \qquad \text{and} \qquad 1 \leq q \leq P. \tag{3.32}$$

The *maximization* step then updates the estimate according to

$$\hat{\mathbf{x}}_P(n+1) \equiv (\mathbf{H}_P^T \mathbf{W}_P(n+1) \mathbf{H}_P)^{-1} \mathbf{H}_P^T \mathbf{W}_P(n+1) \mathbf{R}, \qquad \text{for} \quad n = 0, 1, 2, \ldots \tag{3.33}$$

Theoretically speaking, the inverse of the matrix $\mathbf{H}_P^T\mathbf{W}_P(n+1)\mathbf{H}_P$ should always exist, but numerical issues may arise in its evaluation (see Appendix A).

In essence, the EM algorithm uses the latest estimate to update the weight matrix and uses the new weight matrix to evaluate a new estimate. The likelihood sequence associated with these estimates is monotonically increasing (see Appendix B). Therefore, it is natural to terminate this iterative procedure when the difference of successive likelihoods lies within an acceptable tolerence interval.

3.6 Error Performance

Following the general measures of performance of any unknown estimate, we shall evaluate the bias and the error covariance matrix of our ML parameter-vector estimate, $\hat{\mathbf{x}}_{PML}$.

The error vector associated with $\hat{\mathbf{x}}_{PML}$ is defined by

$$\mathbf{e}_{PML}^{\mathbf{x}} \equiv \mathbf{x}_P - \hat{\mathbf{x}}_{PML}, \tag{3.34}$$

The bias of $\hat{\mathbf{x}}_{PML}$ is then the average error of this estimate,

$$\mathbf{b}_{PML}(\mathbf{X}_P) = E(\mathbf{e}_{PML}^{\mathbf{x}} | \mathbf{x}_P = \mathbf{X}_P), \tag{3.35}$$

where $E(\cdot \mid \mathbf{x}_P = \mathbf{X}_P)$ denotes the expectation with respect to $p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P)$. In general, this value may depend on the true value, $\mathbf{X}_P$, of the parameter vector. Ideally, we would like to have an unbiased estimate which satistifies,

$$\mathbf{b}_{PML}(\mathbf{X}_P) = \mathbf{0}, \qquad \text{for all} \quad \mathbf{X}_P. \tag{3.36}$$

If there are no anomalies, i.e., $\Pr(A) = 0$, the range data vector can be written as

$$\mathbf{r} = \mathbf{H}_P\mathbf{x}_P + \mathbf{v}, \tag{3.37}$$

where $\mathbf{v}$ is a Q-D column vector of independent, identically distributed, Gaussian

random variables each with zero mean and variance δR^2. In this case, using Eq. 3.26 and Eq. 3.38, it can easily be shown that $\mathbf{b}_{PML} = 0$. However, with $\Pr(A) > 0$, the bias cannot be evaluated so explicitly, alhough it has be shown in planar range profiling that there will be a useful CNR regime wherein $\hat{\mathbf{x}}_{PML}$ will be approximately unbiased [12].

The error covariance matrix of the estimate is given by

$$\mathbf{\Lambda}_{PML}(\mathbf{X}_P) \equiv E\left([\mathbf{e}^{\mathbf{x}}_{PML} - \mathbf{b}_{PML}(\mathbf{X}_P)][\mathbf{e}^{\mathbf{x}}_{PML} - \mathbf{b}_{PML}(\mathbf{X}_P)]^T | \mathbf{x}_P = \mathbf{X}_P\right). \quad (3.38)$$

The diagonal elements are the estimation variances for the components of $\hat{\mathbf{x}}_{PML}$, which become the mean-squared estimation errors when the estimate is unbiased,

$$\xi_q^2(\mathbf{X}_P) \equiv E[(x_q - \hat{x}_{qML})^2 | \mathbf{x}_P = \mathbf{X}_P], \qquad \text{for} \quad 1 \le q \le P. \quad (3.39)$$

For *any* unbiased estimate, $\hat{\mathbf{x}}_P$, the Cramér-Rao inequality states that the error covariance matrix should satisfy

$$\mathbf{\Lambda}_P(\mathbf{X}_P) \ge \mathbf{I}_{\mathbf{r}}(\mathbf{X}_P)^{-1}, \quad (3.40)$$

where

$$\mathbf{I}_{\mathbf{r}}(\mathbf{X}_P) \equiv E\left\{\left[\frac{\partial}{\partial \mathbf{x}_P} \ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{x}_P)]\right]\left[\frac{\partial}{\partial \mathbf{x}_P} \ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P)]\right]^T \middle| \mathbf{x}_P = \mathbf{X}_P\right\}, \quad (3.41)$$

is the Fisher information matrix for estimating $\mathbf{x}_P$ from $\mathbf{r}$. In Eq. 3.41, $\frac{\partial}{\partial \mathbf{x}_P}$ operates on a scalar operand producing a P-D column vector whose qth element is the partial derivative of the argument with respect to x_q. The inequality in Eq. 3.40 implies that $\mathbf{\Lambda}_P(\mathbf{X}_P) - \mathbf{I}_{\mathbf{r}}(\mathbf{X}_P)^{-1}$ is a positive semidefinite matrix. Therefore, the diagonal elements of $\mathbf{I}_{\mathbf{r}}(\mathbf{X}_P)^{-1}$ are lower bounds for the corresponding elements of $\mathbf{\Lambda}_P(\mathbf{X}_P)$, i.e., for the mean-squared errors, $\{\xi_q^2\}$, of the unbiased estimator $\hat{\mathbf{x}}_P$. Whenever there is an unbiased estimator whose error covariance matrix equals the inverse of the Fisher

information matrix for all $\mathbf{X}_P$, we call that estimator *efficient.*

In general, for nonlinear estimation problems such as our range profiling case, efficient estimators do *not* exist. But if there is one, it is *always* the ML estimator [17]. An estimator is unbiased and efficient if and only if [18]

$$\hat{\mathbf{x}_P}(\mathbf{R}) = \mathbf{X}_P + \mathbf{I}_\mathbf{r}^{-1}(\mathbf{X}_P)\frac{\partial}{\partial \mathbf{x}_P}\ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{x}_P)]. \tag{3.42}$$

It should be worth noting that the right-hand side of Eq. 3.42 must be independent of $\mathbf{X}_P$ for equality to hold. This allows us to determine the existence of an efficient, unbiased estimator as in our range profiling problem by directly computing the right-hand side of Eq. 3.42. If the result is *not* independent of $\mathbf{X}_P$, we can be certain that an unbiased, efficient estimator cannot exist.

If $\Pr(A) = 0$, the estimation problem is reduced to a Gaussian problem with the ML solution given in Eq. 3.26. This estimate is both unbiased and efficient with an error covariance of

$$\mathbf{\Lambda}_P(\mathbf{X}_P) = \mathbf{I}_\mathbf{r}(\mathbf{X}_P)^{-1} = \delta R^2(\mathbf{H}_P^T\mathbf{H}_P)^{-1} = \delta R^2\mathbf{I}_P. \tag{3.43}$$

With $\Pr(A) > 0$, it is difficult to evaluate the Fisher information matrix due to the log-likelihood's nonlinear dependence on $\mathbf{R}$. Therefore, we will use a weaker, but explicit lower bound for $\mathbf{I}_\mathbf{r}(\mathbf{X}_P)^{-1}$ using the complete data vector $\mathbf{y}$ given in Eq. 3.27 [11]. It is easier to calculate the Fisher information for estimating $\mathbf{x}_P$ from $\mathbf{y}$ since in that case, we are back to a linear problem. Furthermore, any unbiased estimate of $\mathbf{x}_P$ based on $\mathbf{r}$ is *also* an unbiased estimate of $\mathbf{x}_P$ based on $\mathbf{y}$. Thus we find the following complete-data (CD) bound,

$$\mathbf{\Lambda}_P(\mathbf{X}) \geq \mathbf{I}_\mathbf{r}(\mathbf{X}_P)^{-1} \geq \mathbf{I}_\mathbf{y}(\mathbf{X}_P)^{-1} = \frac{\delta R^2}{1 - \Pr(A)}\mathbf{I}_P, \qquad \text{for all} \quad \mathbf{X}_P. \tag{3.44}$$

Note that this CD bound becomes the CR bound when $\Pr(A) = 0$. It has been shown in planar case profiling that the ML estimation performance can approach the

CD bound, even for $\Pr(A) \geq 0$ [11, 12].

3.7 Underresolving the Parameter Vector

Up to this point, our analysis has been based on the assumption that only the first P-dimensions of the parameter vector are nonzero, and that it is these dimensions which we are estimating. In a general range-imaging scenario, we may not know, for a given $\{\phi_q\}$ basis, how many dimensions of $\mathbf{X}$ are nonzero. Indeed, to suppress anomalies, we may underresolve the parameter vector, i.e., estimate $\mathbf{X}_P$ when more than P dimensions of $\mathbf{X}$ are actually nonzero. Consider the case where $\Pr(A) = 0$. The ML estimation problem is reduced to a linear Gaussian problem in which the ML estimate is just the least-squares solution given by Eq. 3.26. If all the components of the full parameter vector are nonzero, their ML estimate is

$$\hat{\mathbf{x}}_{ML}(\mathbf{R}) = \mathbf{H}^T\mathbf{R}. \tag{3.45}$$

Because $\mathbf{H}$ is an orthogonal matrix, the ML estimate of $\mathbf{X}_P$ is just the first P dimensions of this result, viz., we have that

$$\hat{\mathbf{x}}_{ML}(\mathbf{R}) = \begin{bmatrix} \hat{\mathbf{x}}_{PML}(\mathbf{R}) \\ \hat{\mathbf{x}}^c_{PML}(\mathbf{R}) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_P^T & \mathbf{H}_P^{cT} \end{bmatrix} \mathbf{R}. \tag{3.46}$$

With $\Pr(A) > 0$, however, the situation is a bit more complicated. Owing to the orthogonality of the transformation basis and the diagonality of the weight matrix, we can perform EM iterations separately on the P dimensions that we are interested in. Unlike the case with $\Pr(A) = 0$, estimating with P dimensions will not give the same P components as we would get from the full-parameter estimation due to the presence of anomalies. This will be problematic if P is too small, because a lot of the pixels will be treated as anomalous and discarded by the EM algorithm, resulting in a loss of fine-scale features that might be of interest. These underresolved estimates tend to be biased with significant RMS errors. Examples of underresolved profiles will be shown later in Sec. 4.5.

Chapter 4

Multiresolution Range Profiling

The purpose of employing a parametric ML range profiler is to impose regularity conditions to ensure a degree of anomaly suppression. In general scenarios, the choice of a parametric model, i.e., an $\{\mathbf{H}, P\}$ pair is not very clear.

Suppose we arrange the $\{\phi_q\}$ such that increasing q corresponds to increasingly fine scale behavior of some sort. It follows naturally that we would want to extract the coarse-to-fine features of the range-data by successively increasing the value of P at each stage of the estimation process. The estimation process should terminate when we reach the finest scale of interest or when P become so large that we are unsure if the anomaly suppression is sufficient to warrant any finer-scale estimate. This is where the use of wavelet basis [15] and multiresolution signal processing [13, 14] can be incorporated.

The main focus of this thesis is to develop a multiresolution approach for laser radar range profiling. We will use a Haar wavelet as our orthonormal basis for the discussion and simulation from here onwards, although the general construct we develop is applicable to any orthonormal wavelet basis.

4.1 Multiresolution Wavelet Basis

4.1.1 One-dimensional Haar Wavelets

To obtain some initial understanding of the behavior of the algorithm, the range-estimation problem is reduced to fitting a 1-D profile to a single elevation row of a 2-D range image. Thus we assume $Q = K$ and $\{r_q \equiv r_{j_o q} : 1 \leq q \leq Q\}$, for some particular j_o. Furthermore, we assume $Q = 2M$, where M is an integer, and we have

$$\boldsymbol{\phi}_q = \begin{bmatrix} \phi_{q1} \\ \phi_{q2} \\ \vdots \\ \phi_{qQ} \end{bmatrix}, \qquad \text{for} \quad 1 \leq q \leq Q, \tag{4.1}$$

where

$$\phi_{1n} = \frac{1}{\sqrt{Q}}, \qquad \text{for} \quad 1 \leq n \leq Q, \tag{4.2}$$

and

$$\phi_{qn} = \frac{\psi_p[n[2^{-(p-1)}(q-1)-1]Q]}{\sqrt{Q}},$$
$$\text{for} \quad 1 \leq n \leq Q \quad \text{and} \quad 1 \leq 2^{p-1} < q \leq 2^p \leq Q. \tag{4.3}$$

In Eq. 4.3,

$$\psi_p[n] \equiv 2^{(p-1)/2}\psi[2^{p-1}n], \qquad \text{for} \quad 1 \leq n \leq Q, \tag{4.4}$$

where

$$\psi[n] \equiv \begin{cases} 0 & : \quad n \leq 0, \\ 1 & : \quad 1 \leq n \leq Q/2, \\ -1 & : \quad Q/2 < n \leq Q, \\ 0 & : \quad n > Q. \end{cases} \tag{4.5}$$

To illustrate the nature of the Haar wavelet, below is an example for $Q = 8$. Using Eqs. 4.2 - 4.5, we have

$$\boldsymbol{\phi}_1^T = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \tag{4.6}$$

$$\boldsymbol{\phi}_2^T = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}, \tag{4.7}$$

$$\boldsymbol{\phi}_3^T = \frac{1}{2} \begin{bmatrix} 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{4.8}$$

$$\boldsymbol{\phi}_4^T = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix}, \tag{4.9}$$

$$\boldsymbol{\phi}_5^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{4.10}$$

$$\boldsymbol{\phi}_6^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{4.11}$$

$$\boldsymbol{\phi}_7^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}, \tag{4.12}$$

$$\boldsymbol{\phi}_8^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}. \tag{4.13}$$

The wavelets are shown in Fig. 4-1.

It is not difficult to verify that the $\{\boldsymbol{\phi}_q\}$ basis constructed is orthonormal. More importantly, the self-similar nature of the wavelets demonstrates a progression of finer-scale behavior as q increases. Thus $\{\boldsymbol{\phi}_q : 1 \leq q \leq P \equiv 2^p < Q\}$ spans a set of 1-D range profiles, $\{r_q : 1 \leq q \leq Q\}$, for which r_q is piecewise constant over Q/P-length intervals.

Using a $\{\mathbf{H}, P\}$ pair in the ML estimation scheme, where $P < Q$, we should be able to fit a piecewise constant profile to the range data with strong anomaly suppression. The multiresolution nature of the Haar-wavelet basis allows us to sequence through a progression of increasing values of P which corresponds to increasingly fine piecewise-

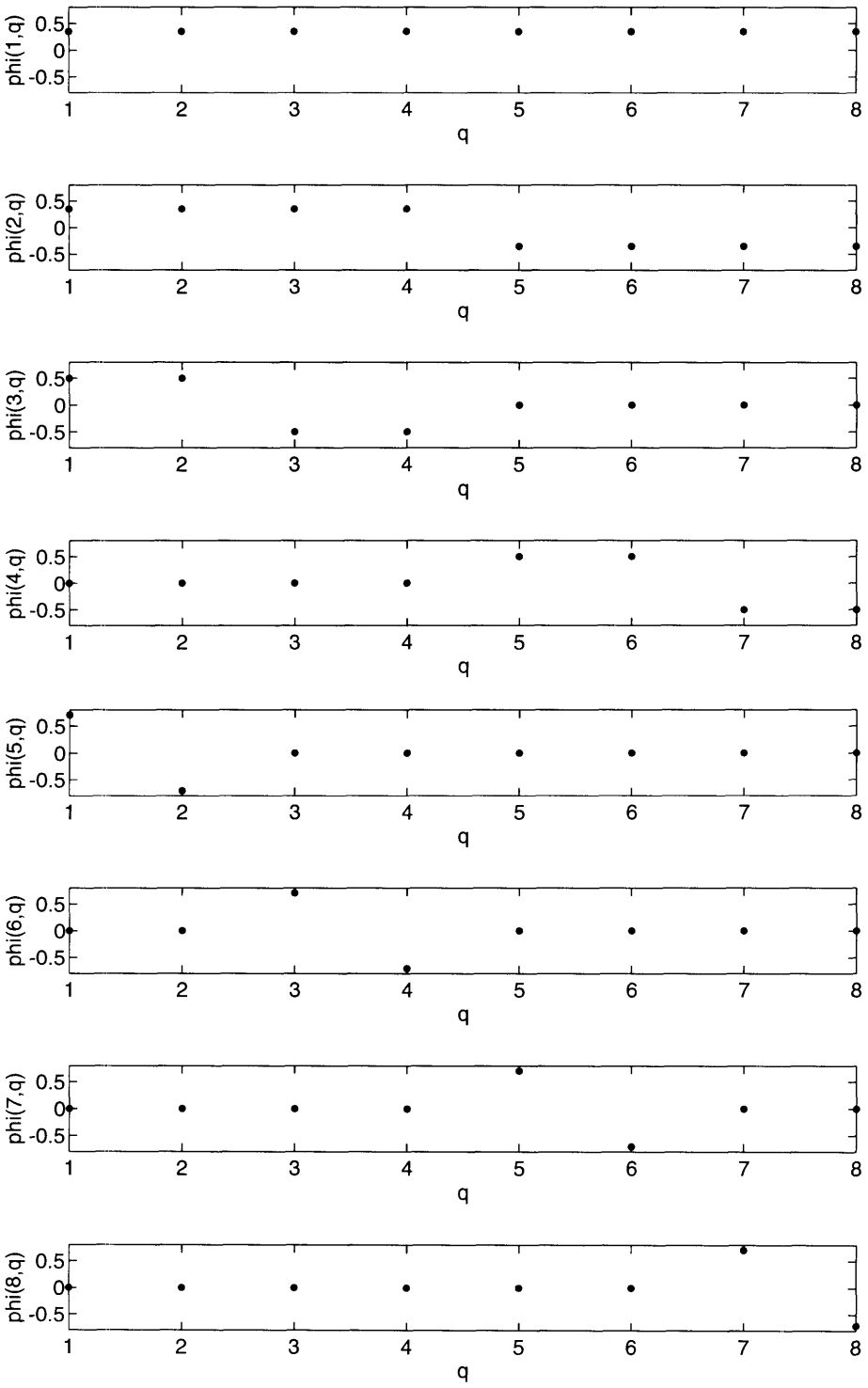

Figure 4-1: A Haar-wavelet $\{\phi_q\}$ for $Q = 8$.

approximations to the true-range profile.

4.1.2 Two-dimensional Haar Wavelets

The extension of the one-dimensional Haar wavelet basis to a two-dimensional setting is quite straightforward. A simple way of initial extension is to employ the separable Haar basis generated by the product of two 1-D bases.

Suppose we let $\mathbf{H}_J$ and $\mathbf{H}_K$ be the two initial 1-D bases, that is

$$\mathbf{H}_J \equiv [\ \boldsymbol{\phi}_1 \ \cdots \ \boldsymbol{\phi}_j \ \cdots \ \boldsymbol{\phi}_J\ ] \quad \text{for} \quad 1 \leq j \leq J, \tag{4.14}$$

where

$$\boldsymbol{\phi}_j = \begin{bmatrix} \phi_{j1} \\ \phi_{j2} \\ \vdots \\ \phi_{jQ} \end{bmatrix}, \quad \text{for} \quad 1 \leq j \leq J, \tag{4.15}$$

and

$$\mathbf{H}_K \equiv [\ \boldsymbol{\phi}_1 \ \cdots \ \boldsymbol{\phi}_k \ \cdots \ \boldsymbol{\phi}_K\ ] \quad \text{for} \quad 1 \leq k \leq K, \tag{4.16}$$

where

$$\boldsymbol{\phi}_k = \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kQ} \end{bmatrix}, \quad \text{for} \quad 1 \leq k \leq K. \tag{4.17}$$

be the $Q \times J$ and $Q \times K$ 1-D Haar-wavelet coordinate transformations. For consistency with the 1-D basis described in the previous section, we shall assume that $J = 2^M$ and $K = 2^N$ where M and N are both integers.

Let $\{\boldsymbol{\zeta}_{jk} : 1 \leq j \leq J, 1 \leq k \leq K, Q \equiv JK\}$ be the separable column-vector basis for the Q-D vector space, and let

$$\mathbf{H}_{J\times K} \equiv [\ \boldsymbol{\zeta}_{11} \ \ \boldsymbol{\zeta}_{12} \ \cdots \ \boldsymbol{\zeta}_{JK}\ ] \tag{4.18}$$

be the $Q \times Q$ coordinate transformation matrix constructed from the 1-D bases $\{\phi_j : 1 \leq j \leq J\}$ and $\{\phi_k : 1 \leq k \leq K\}$ where

$$\zeta_{jk} = \begin{bmatrix} \zeta_{jk1} \\ \zeta_{jk2} \\ \vdots \\ \zeta_{jkQ} \end{bmatrix} = \begin{bmatrix} \phi_j \phi_{k1} \\ \phi_j \phi_{k2} \\ \vdots \\ \phi_j \phi_{kQ} \end{bmatrix}. \tag{4.19}$$

Notice that we can associate a different value of P for the two independent 1-D bases. This means that we can estimate the range data using different resolutions for the J and K directions. For notational clarity, let us define $P_j \equiv 2^j$ to be the number of nonzero parameters to be estimated in the J direction and $P_k \equiv 2^k$ to be the number in the K direction. Thus we have

$$\mathbf{H}_{P_j P_k} \equiv [\ \zeta_{11} \quad \zeta_{12} \quad \cdots \quad \zeta_{P_j P_k}\ ] \tag{4.20}$$

The structure of the 2-D separable Haar wavelet basis so constructed is the same as the 1-D basis except for the increase in dimensionality and the extra freedom in the choice of P_j and P_k. Therefore, using the same algorithm as discussed in the previous chapter, we shall be able to obtain an ML estimate for the range profile.

In general, there is no need to restrict the 2-D basis to be separable. In many scenarios, *non*separable wavelet bases may be more useful. The best choice of a multiresolution basis may be application-dependent and is still an open research problem.

4.2 Simulated Range Data

Computer-simulated range images are generated based on the statistical model presented in Sec. 2.1. An arbitrary range truth is chosen which resembles a "Manhattan skyline". Fig. 4-2 shows the 1-D 512-pixel range truth used in the testing of the ML estimation algorithm. The horizontal scale represents the pixels and the vertical scales is in meters. The range uncertainty interval (ΔR) assumed for the laser radar

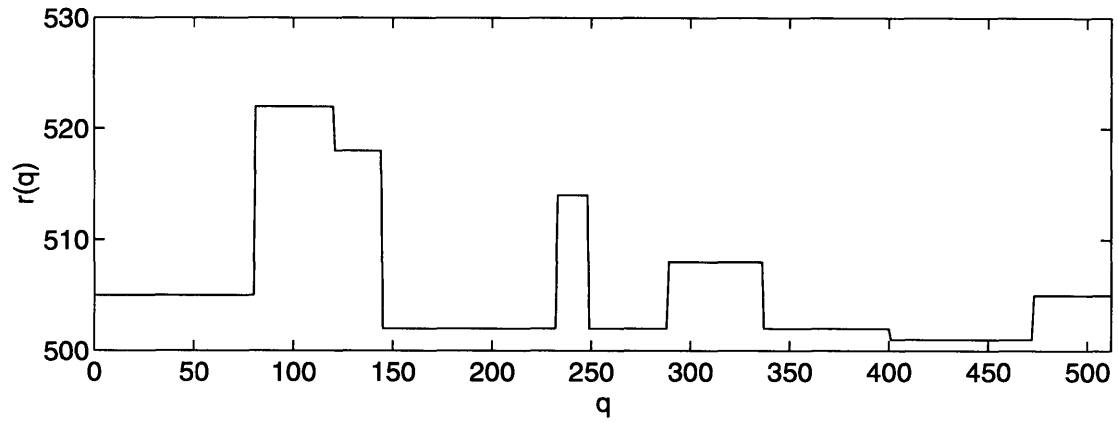

Figure 4-2: A range truth of 512 pixels.

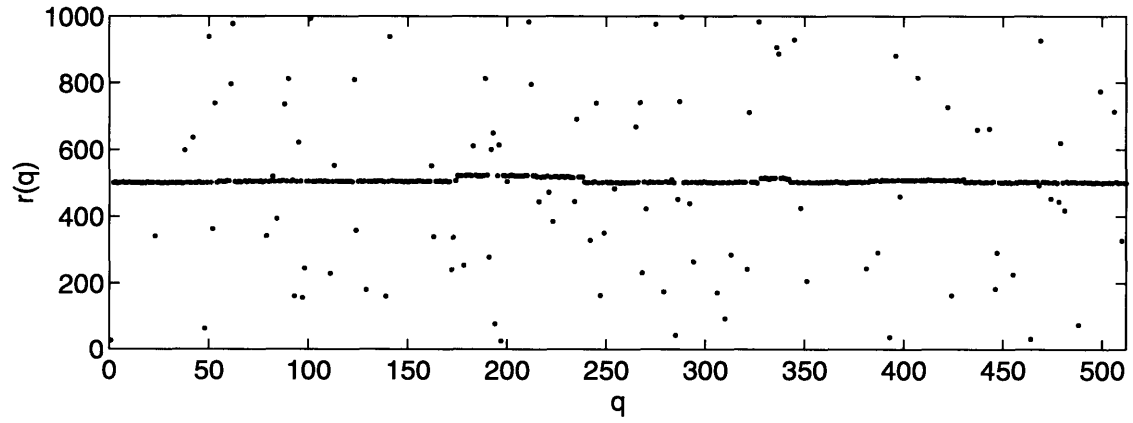

Figure 4-3: Typical simulated range data with anomaly probability of 0.2.

is 1000 m and the local range accuracy (δR) is 1 m. Fig. 4-3 shows typical range data with $\Pr(A) = 0.2$. The data points close to the solid line are the correct data points while the anomalous data points are scattered all over the uncertainty interval.

For 2-D range images, the simplest approach is to use the product of two range truths of length N to generate a $N \times N$ true range. However, the image thus constructed will only contain features that are aligned with the azimuth and the elevation coordinates, as shown in Fig. 4-4.

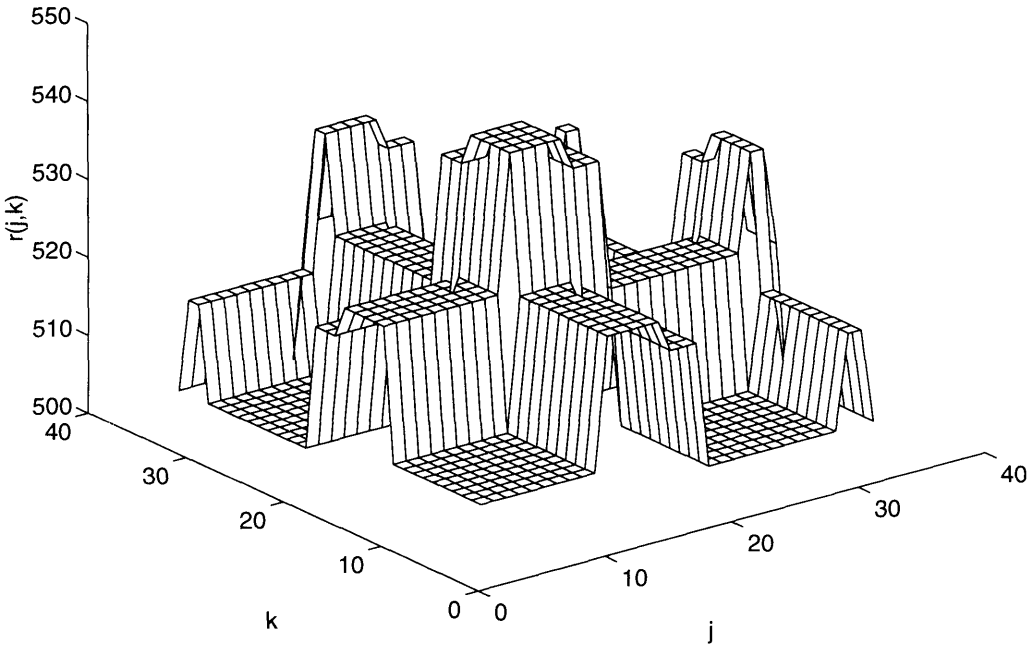

Figure 4-4: A 2-D range truth of 32×32 pixels.

4.3 Initialization Methods

As discussed earlier in Sec. 3.5, the EM algorithm will converge to a local likelihood maximum through generating a sequence of estimates with increasing likelihoods. In theory, an arbitrary initialization can be used, although whether or not the EM algorithm converges to the global likelihood maximum depends on the quality of this initial estimate. We will discuss two initialization methods in the following subsections: their resulting estimation performances will be compared in Sec. 4.5.

4.3.1 Range-Truth Initialization

With simulated data, the true parameter vector, $\mathbf{X}_P$, is available for a given range truth. Thus we can approximate the ML estimate via the EM algorithm by setting $\hat{\mathbf{x}}_P(0) = \mathbf{X}_P$ and iterate until a stationary point is reached. Initializing the EM algorithm with the true parameter will result in an estimate whose likelihood is even higher that of the true parameter. Thus we can expect this estimate to be very near to $\hat{\mathbf{x}}_{PML}$. However, there is a problem of underresolving the range-truth when fewer parameters are estimated than are implied by the original profile, as discussed in Sec. 3.7. Moreover, this initialization scheme is not possible for real laser radar range profiles as the range-truth is unknown. In that case, more robust initialization techniques are needed [12].

Using the range truth, we obtain our initialization:

$$\hat{\mathbf{x}}_P(0) = \mathbf{H}_P^T \mathbf{r}^*. \tag{4.21}$$

For notational simplicity, we will use the value of P to denote resolution from here onwards such that for a particular P, the smallest piecewise-constant interval that can be represented is of length Q/P. In other words, higher P corresponds to higher resolution.

Figs. 4-5 – 4-8 show the initial estimates supplied to the EM algorithm at different resolutions, where the dash-dotted line is the range truth. These seeds are just the Haar-wavelet representations of the range truth for different values of P. Notice that

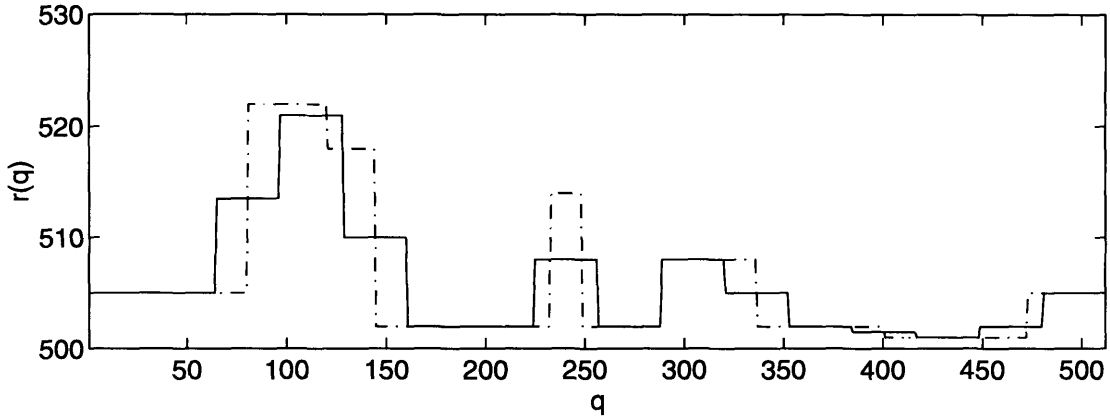

Figure 4-5: Haar-fitted profile at $P = 16$ for initializing the EM algorithm; dash-dotted line is the range-truth.

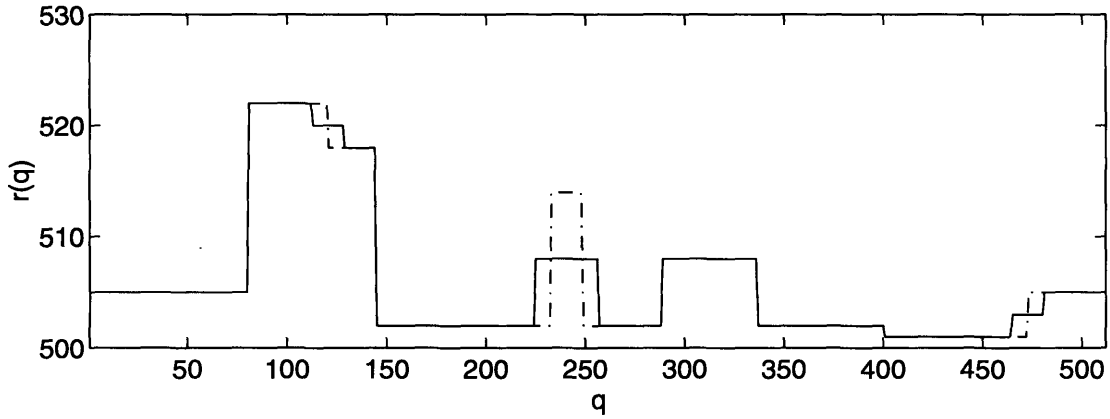

Figure 4-6: Haar-fitted profile at $P = 32$ for initializing the EM algorithm; dash-dotted line is the range-truth.

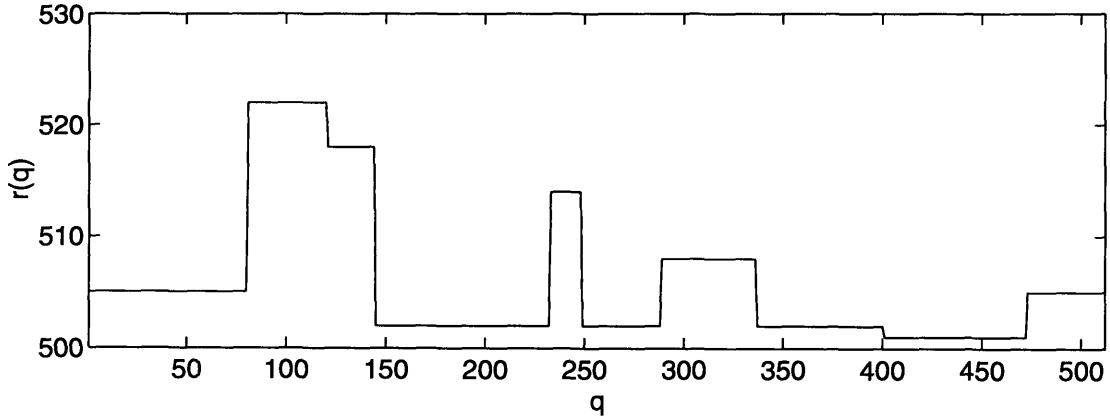

Figure 4-7: Haar-fitted profile at $P = 64$ for initializing the EM algorithm; the range truth is identical to the $P - 64$ Haar fit.

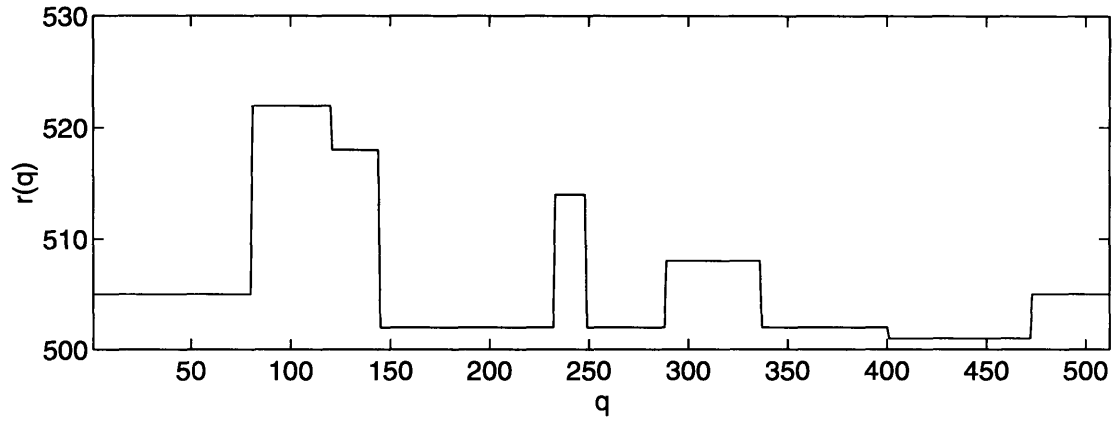

Figure 4-8: Haar-fitted profile at $P = 128$ for initializing the EM algorithm; the range truth is identical to the $P = 128$ Haar fit.

at low resolution the Haar-fitted profiles tend to average the effects of finer details like narrow peaks and jumps. As the resolution increases, the fitted profiles more closely resemble the range truth. The true parameter vector $\mathbf{X}$ for this range truth has 25 nonzero parameters, which require 64 Haar-wavelets for a perfect representation.

Figs. 4-9 - 4-12 shows the estimated profiles obtained from the EM algorithm at different resolutions. Again, the dash-dotted line is the range truth. It is obvious that at low resolution the finer peaks are treated as anomalous pixels, and are rejected by the EM algorithm (Figs. 4-9 – 4-10). As the resolution increases, more details are captured by the EM algorithm, resulting in an estimated profile that better resembles the range-truth (Fig. 4-11 – 4-12). Due to the nature of the Haar-wavelets, the finer features may be shifted or widened. In addition, the finer-scale estimate also picks up the local noise at the "wide-flat" areas. These undesirable fine details lead to the suggestion of a smoothing scheme for the final estimate which will be discussed in Chap. 6.

4.3.2 Least-Squares Initialization

When the true parameter vector $\mathbf{X}_P$ is not available, as in the case of real laser radar images, a more realistic initialization approach would be the least-squares (LS) initialization as described by Eq. 3.30.

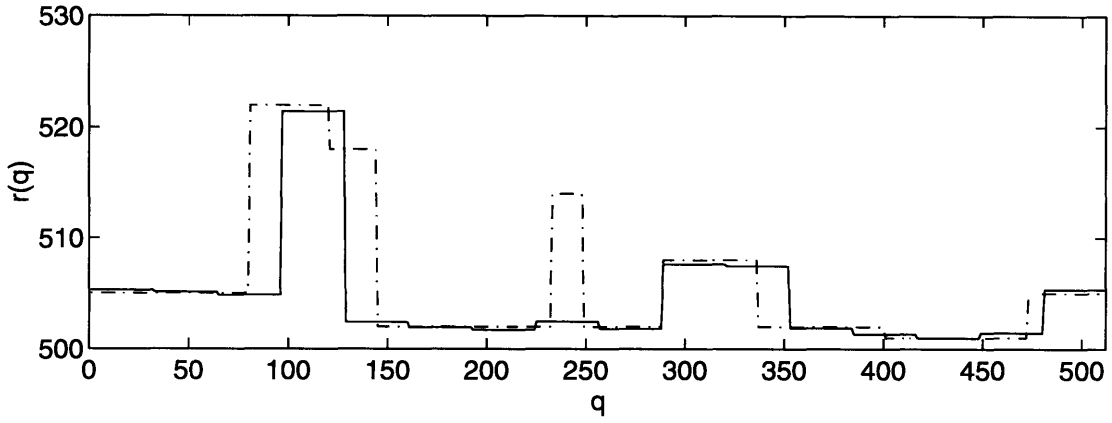

Figure 4-9: Estimated profile seeded with range-truth at $P = 16$; dash-dotted line is the range-truth.

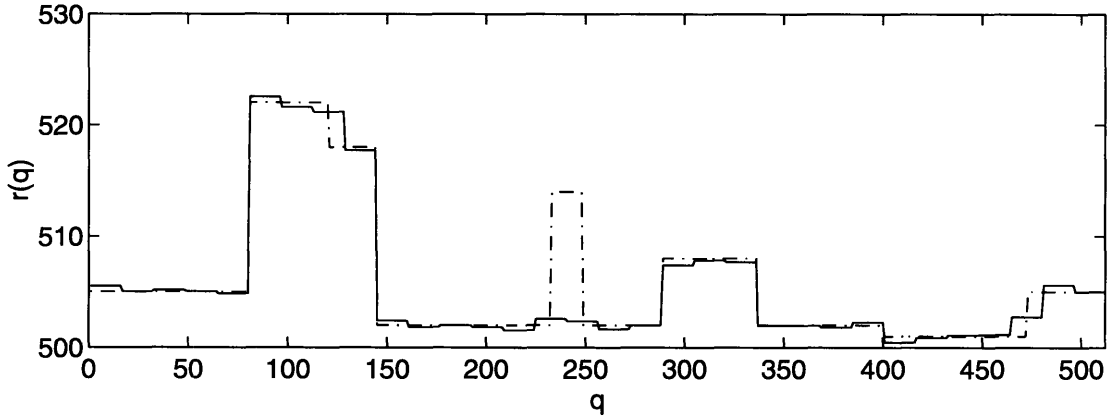

Figure 4-10: Estimated profile seeded with range-truth at $P = 32$; dash-dotted line is the range-truth.

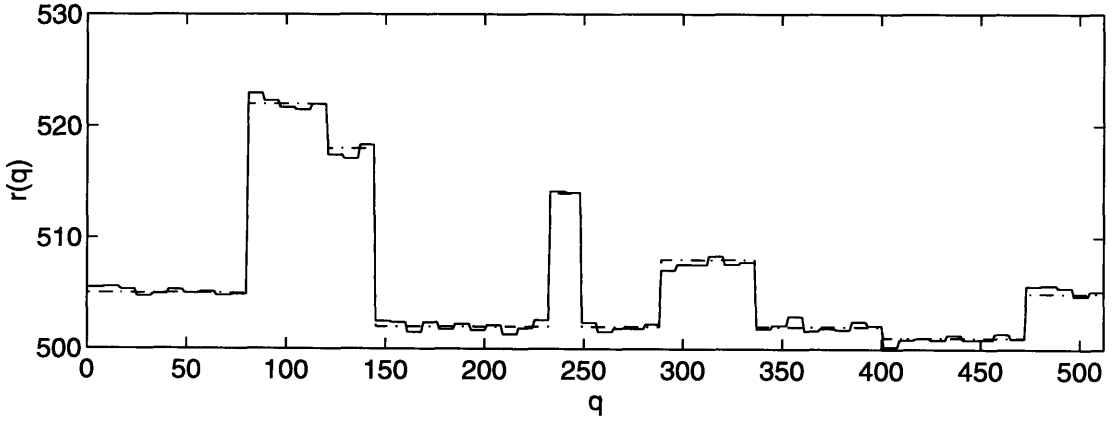

Figure 4-11: Estimated profile seeded with range-truth at $P = 64$; dash-dotted line is the range-truth.

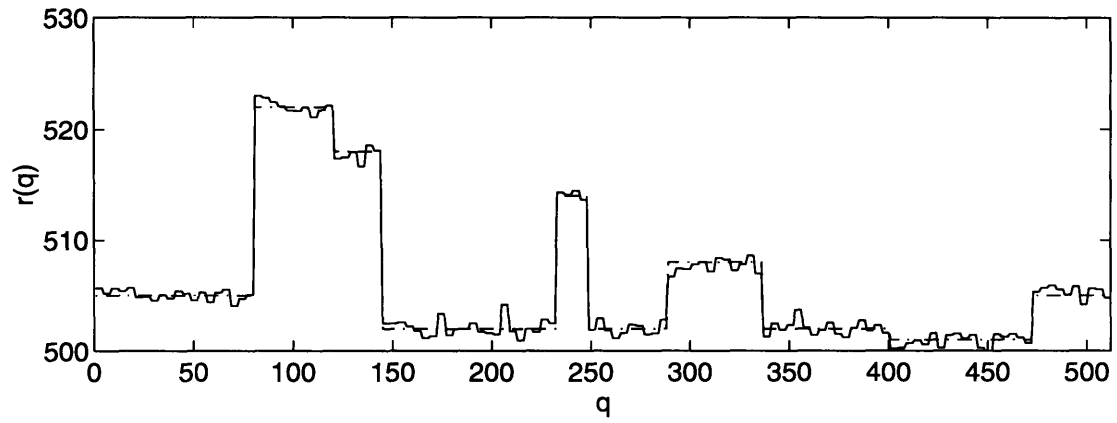

Figure 4-12: Estimated profile seeded with range-truth at $P = 128$; dash-dotted line is the range-truth.

In planar-range profiling case, the LS-initialized EM algorithm finds the ML estimate reliably for large images and low $\Pr(A)$ [12]. The fact that only three parameters – the azimuth and the elevation angles, and the range-intercept – are fitted with a large number of pixels makes the impact of anomalies on the estimate relatively weak. However, as $\delta R/\Delta R$ decreases, only a few anomalous pixels are needed to place the estimated plane far away from the truth and least-squares may not reliably place the initial estimate on the slope where the global maximum resides.

Unfortunately, in our case of multiresolution range profiling, this simple initialization method is in general not sufficient to start the EM algorithm. For small values of P, we again encounter the problem of underresolution where the initial estimated profile is placed many δR away from the true profile. Initial LS estimates at different resolutions for a particular range data are shown in Figs. 4-13 – 4-16, where the dash-dotted line is again the range truth. Due to the nature of the least-squares method, these initial estimates average all the pixels. As a result, a large portion of the nonanomalous pixels are treated as anomalies and are thus discarded in the EM iterations. The final estimates thus obtained (which are not shown) will not be very good approximations of the ML estimates. Indeed, quite often, the weights in Eq. 3.33 are so small, when LS initialization is employed, that the matrix $\mathbf{H}_P^T\mathbf{W}_P(n+1)\mathbf{H}_P$ becomes too singular to invert numerically.

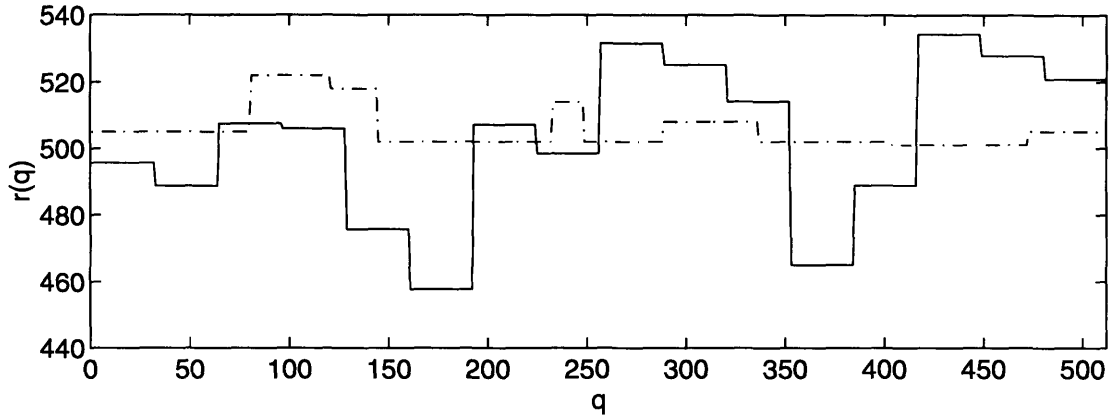

Figure 4-13: Initial least-squares estimated profile at $P = 16$; dash-dotted line is the range-truth.

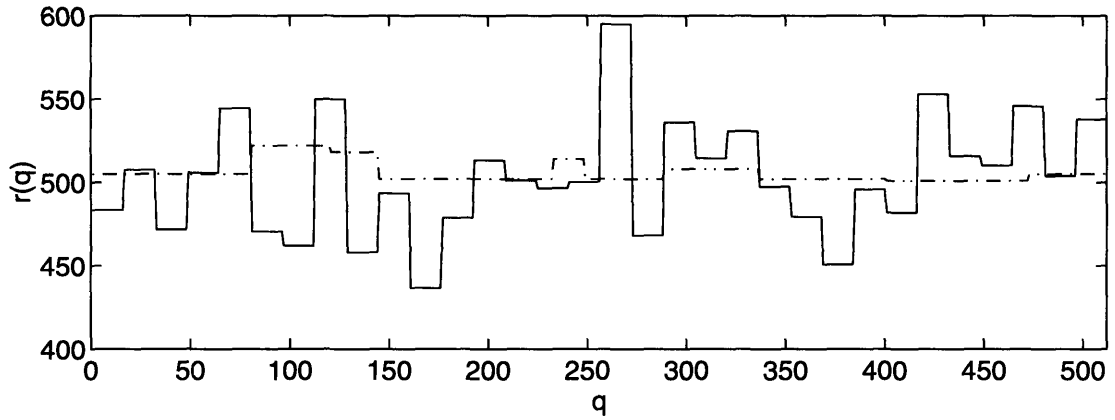

Figure 4-14: Initial least-squares estimated profile at $P = 32$; dash-dotted line is the range-truth.

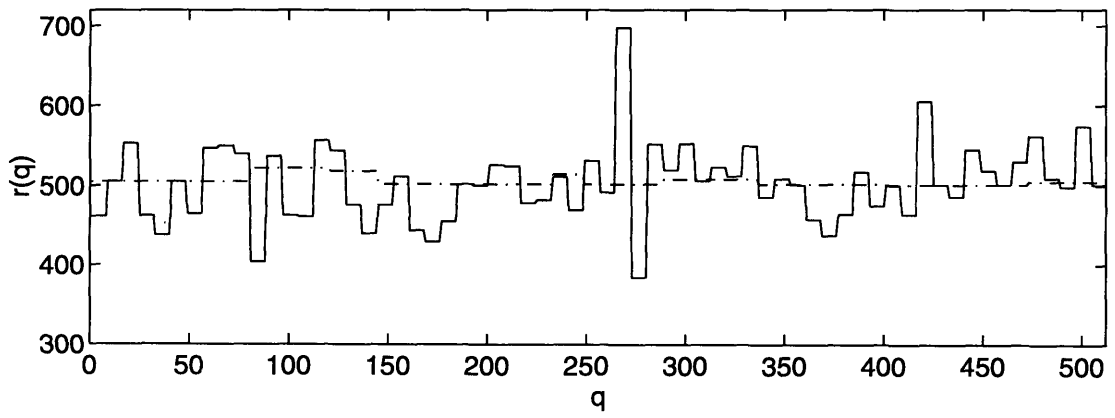

Figure 4-15: Initial least-squares estimated profile at $P = 64$; dash-dotted line is the range-truth.

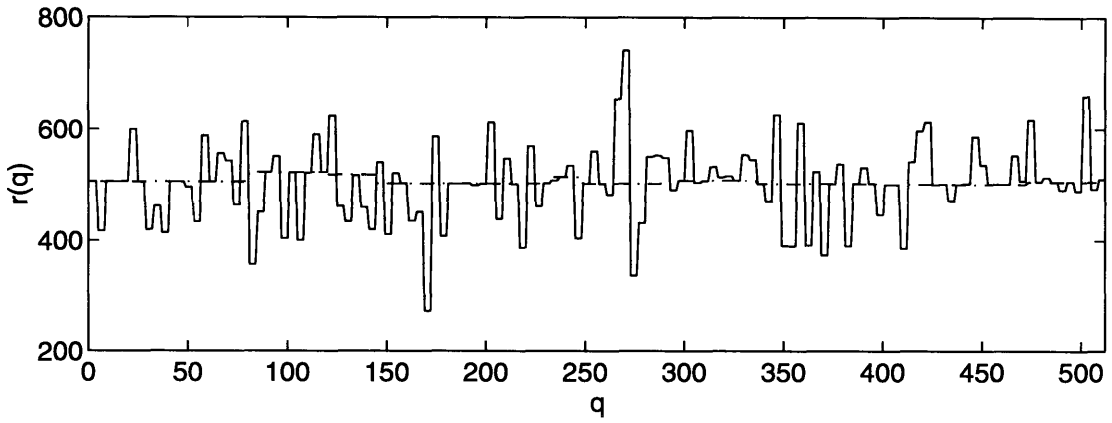

Figure 4-16: Least-squares initial estimated profile at $P = 128$; dash-dotted line is the range-truth.

An alternative initialization approach, the recursive EM (REM) [11] will be described in the next section.

4.4 Recursive EM Algorithm

The REM algorithm is an extension of the least-squares initialized EM estimation. It has been demonstrated in the planar profiling work that the REM algorithm obtains a more reliable ML estimate than the LS-initiated EM algorithm [12].

Suppose we let M be an integer satisfying

$$M \approx \log_2\left(\frac{\Delta R}{\delta R}\right). \tag{4.22}$$

The REM algorithm begins by setting the local range accuracy, δR, in Eq. 2.1 to be $\delta r_0 \equiv \Delta R$. The LS-initialization is then employed on the resulting density to obtain an initial estimate $\hat{\mathbf{x}}_{REM}(0)$ via the EM iterations. The REM algorithm then resets the local range accuracy in Eq. 2.1 to $\delta r_1 \equiv \delta r_0/2 = \Delta R/2$. The EM algorithm is initiated again using $\hat{\mathbf{x}}_{REM}(0)$ as the seed to obtain the first-order REM estimate, $\hat{\mathbf{x}}_{REM}(1)$. This recursive process will repeat until $\hat{\mathbf{x}}_{REM}(M-1)$ is used to initiate the EM iterations with $\delta r_M \equiv \delta r_{M-1}/2 = 2^{-M}\Delta R = \delta R$ used as the local range accuracy. The output of this step is the final REM estimate $\hat{\mathbf{x}}_{REM} = \hat{\mathbf{x}}_{REM}(M)$.

The REM process described above has significant improvements over the LS-initialized EM algorithm. Starting from a very large value (the range uncertainty interval ΔR), the local range accuracy in Eq. 2.1 is reduced by a factor of 2 during each round of the EM-iterations. As a result, only a few pixels are treated as anomalies and discarded at each round. In contrast to the LS-initialized EM algorithm, the REM reduces the chance that a lot of the nonanomalous pixels are disqualified from the estimate because of the quality of the LS initial estimate. Therefore, it is more likely that the REM estimate will be closer to the ML estimate that we desire. Figs. 4-17 – 4-19 show the estimated profiles obtained using the REM algorithm for different resolutions.

4.5 Performance

In this section, we will examine the estimation performance of the multiresolution range profiler via its normalized bias and RMS error. All simulations assumed the same range truth shown in Fig. 4-2. The original profile is best-fitted at a resolution of $P = 64$, in which 25 of the true parameters are nonzero. In the simulation figures, each data point is the sample mean, and the associated error bars are plus-minus one sample standard deviation, of 500 trials.

Figs. 4-20 – 4-22 show the normalized bias of $\hat{\mathbf{x}}_{PML}$, i. e. $(\mathbf{b}_{PML}(\mathbf{X}_P))_q/\delta R$ vs. q,

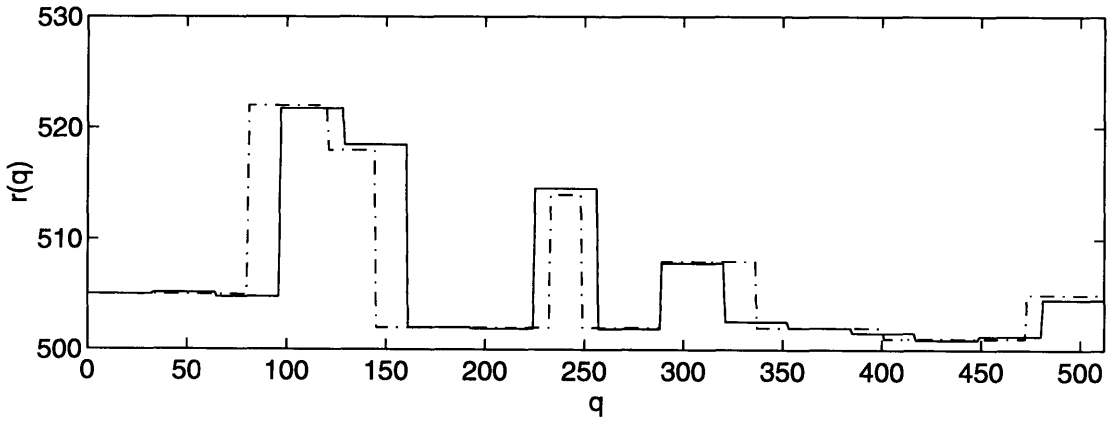

Figure 4-17: REM estimated profile at $P = 16$; dash-dotted line is the range-truth.

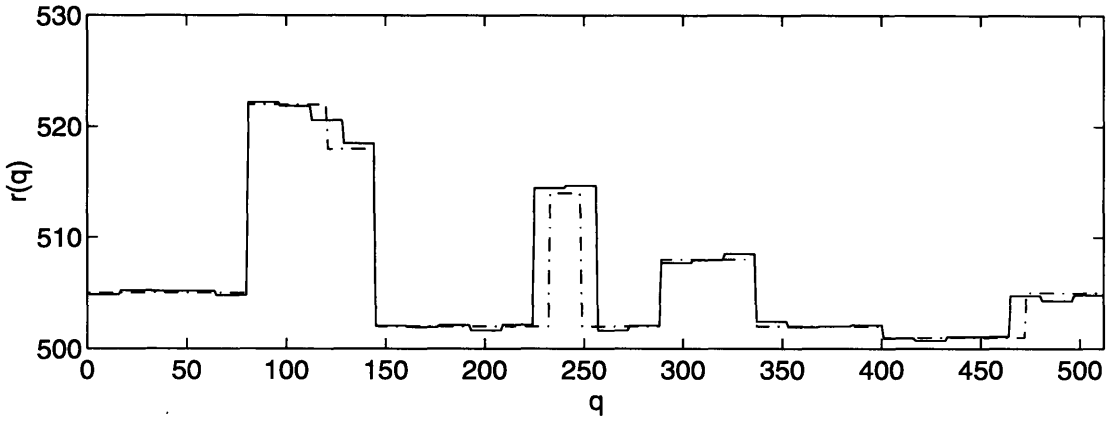

Figure 4-18: REM estimated profile $P = 32$; dash-dotted line is the range-truth.

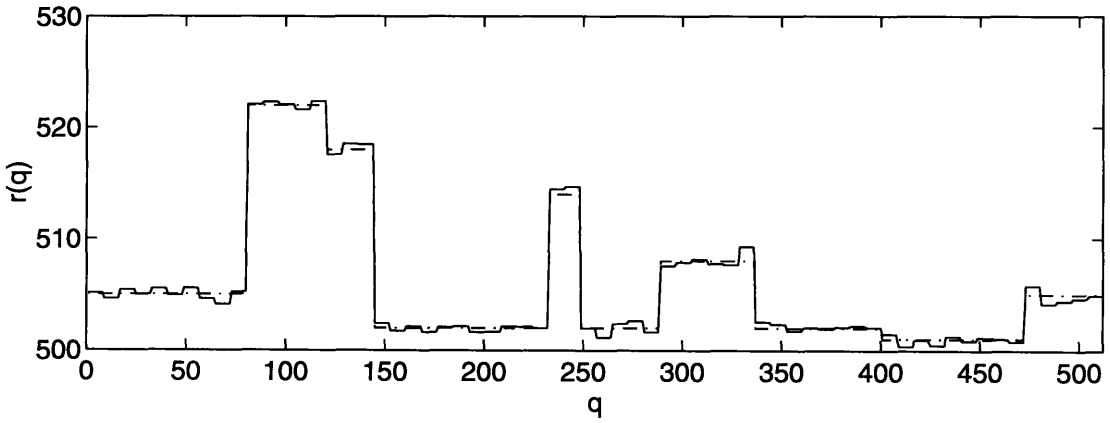

Figure 4-19: REM estimated profile $P = 64$; dash-dotted line is the range-truth.

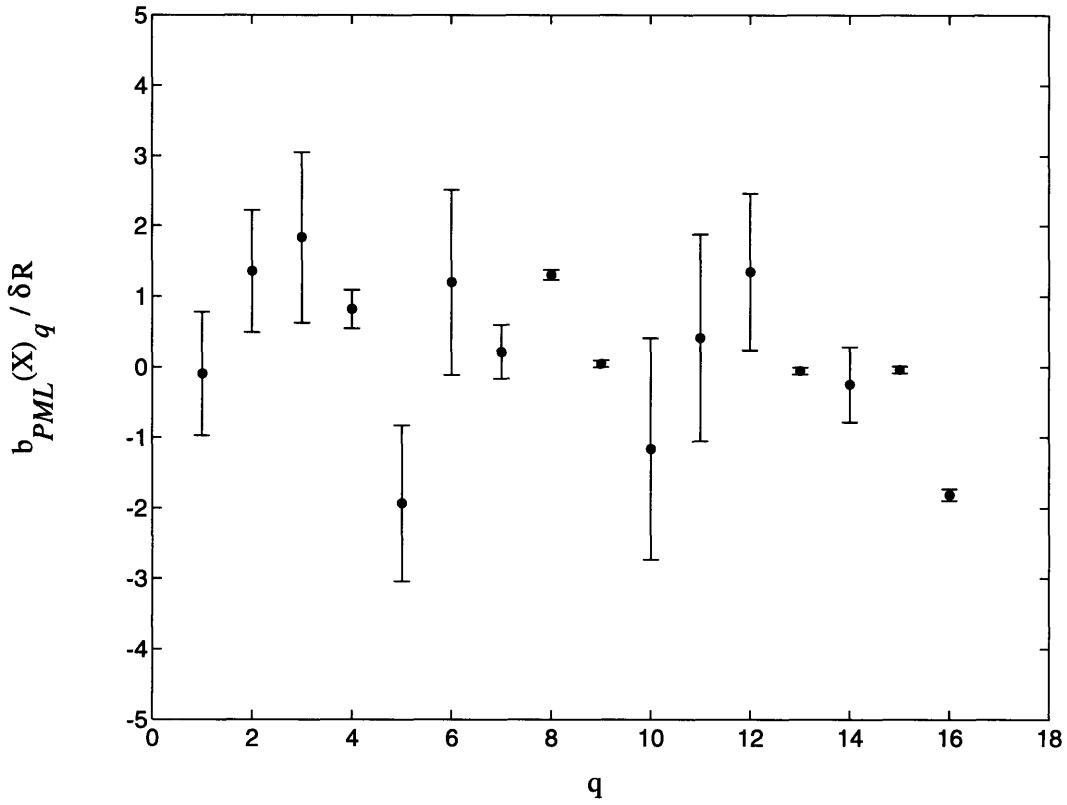

Figure 4-20: Normalized bias of the EM parameter estimates at $P = 16$

where $1 \leq q \leq P$, for different resolutions. We can see that at low resolutions, say $P = 16$, the ML estimate seems to be quite biased. This can be easily understood since we are only using a small number of parameters in the estimation and in that case we are underresolving the range truth. This results in a loss of fine-scale features and we should well-expect the estimate to be biased. As the value of P increases, the estimate becomes less biased. The stopping rule for the estimation process, which will be discussed later in Chap. 5, indicated that the best estimate is at $P = 64$. This corresponds to the resolution of the original profile used, and we can see that our ML estimate is approximately unbiased at that resolution.

Apart from the bias, we are also interested in the RMS estimation error in comparing the ML estimation with the complete-data bound. Figs. 4-23 - 4-25 show the normalized RMS errors, $\xi_{PML_q}(\mathbf{X})/\xi_{CD_q}(\mathbf{X})$ vs. q, at different resolutions. Similar to the behavior of the normalized bias, the RMS errors decrease as more parameters

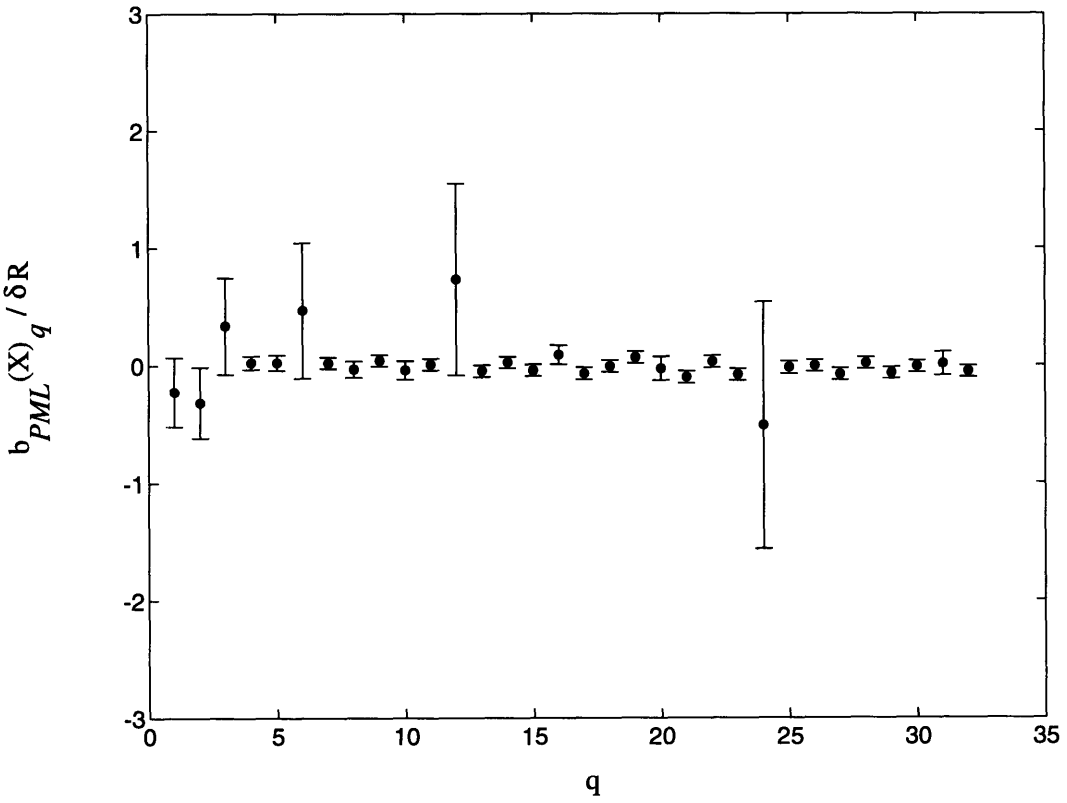

Figure 4-21: Normalized bias of the EM parameter estimates at $P = 32$

are being used in the estimation. In particular, we have $\xi_{PML_q}(\mathbf{X}) \approx \xi_{CD_q}(\mathbf{X})$ at the terminating resolution. Thus for this particular range truth, initializing with the true parameter vector does ensure convergence to the ML estimate at $\Pr(A) = 0.2$.

In general, we do not want to use resolutions higher than a quarter of the full resolution. One reason is that it is numerically difficult to solve Eq. 3.33 due to matrix singularity. Furthermore, the anomaly suppression is weaker as resolution increases; with overly high resolution, anomalous pixels may masquerade as high-resolution features.

As discussed in Sec. 4.3.2, initializing the EM algorithm using the least-squares estimate is seldom reliable. Therefore, we resort to using the REM algorithm as a practical realization of finding the ML estimate. Figs. 4-26 – 4-26 plot the normalized bias of the $\hat{\mathbf{x}}_{PML}$ estimates at different resolutions. The corresponding normalized RMS errors are shown in Figs. 4-29 – 4-31. Here we see behavior similar to that

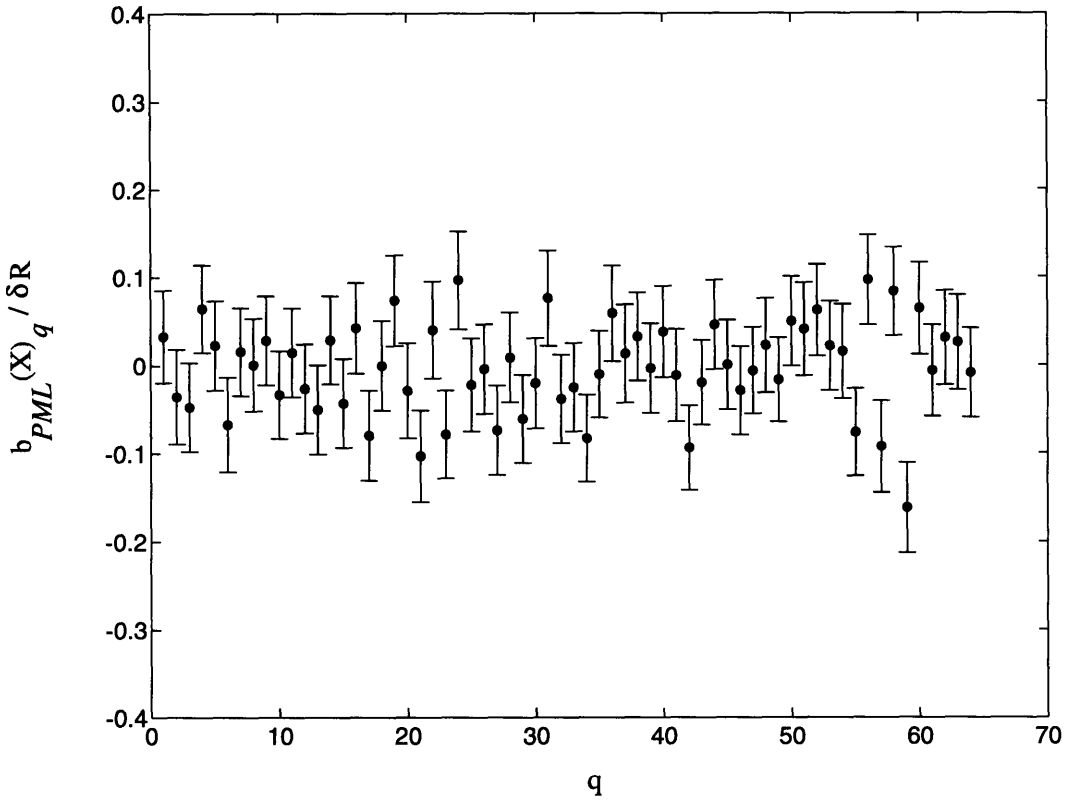

Figure 4-22: Normalized bias of the EM parameter estimates at $P = 64$

seen earlier with the range-truth initialized estimate, with one exception. At high resolution, the RMS errors of the REM algorithm is much worse. This is primarily due to the inclusion of anomalous data points in the first few cycles of the recursive process when the algorithm's local range accuracy is still very coarse, resulting in the inclusion of "false" fine-scale features.

For the 2-D simulations, samples of the estimated profile at different resolutions are shown in Figs. 4-32 – 4-34. The range truth itself, as shown in Fig. 4-4, is separable, making the separable 2-D Haar basis a potentially good coordinate transformation. Basically, the behavior of the system in 2-D profiling is very similar to the 1-D case, despite the fact that the computational time goes up as n^2. However, for images that contain features that are not oriented along the axes, the Haar basis may not be as suitable.

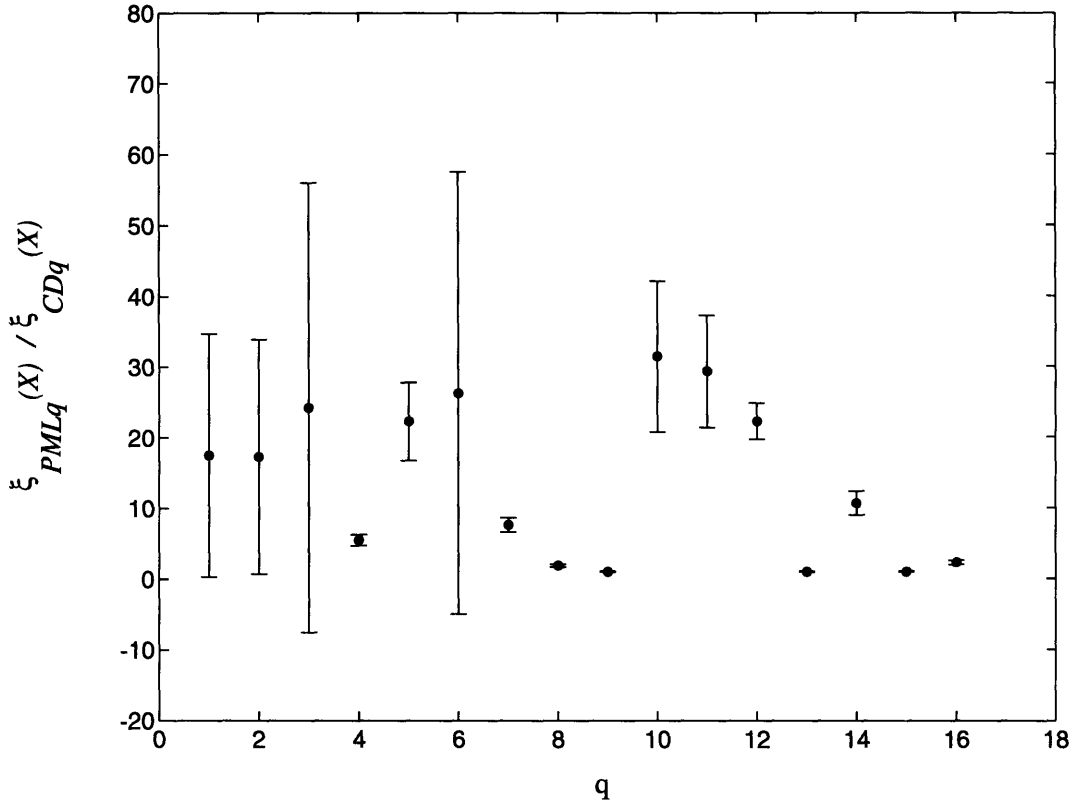

Figure 4-23: Normalized RMS errors of the EM parameter estimates at $P = 16$

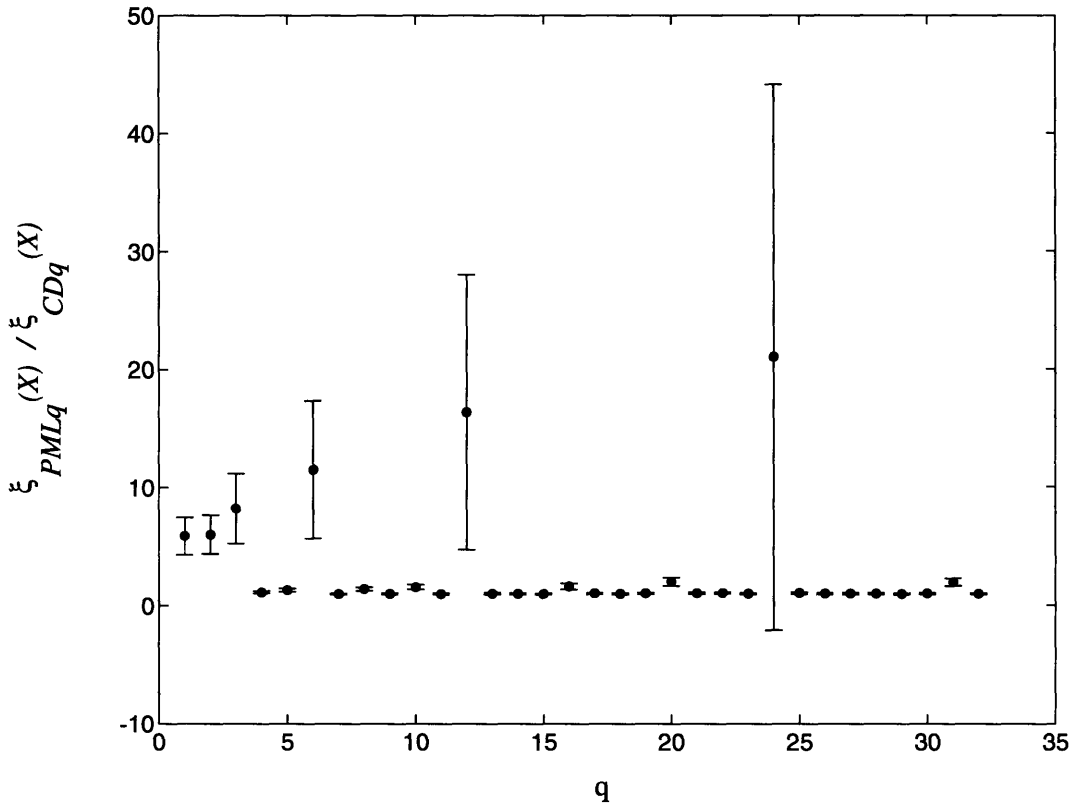

Figure 4-24: Normalized RMS errors of the EM parameter estimates at $P = 32$

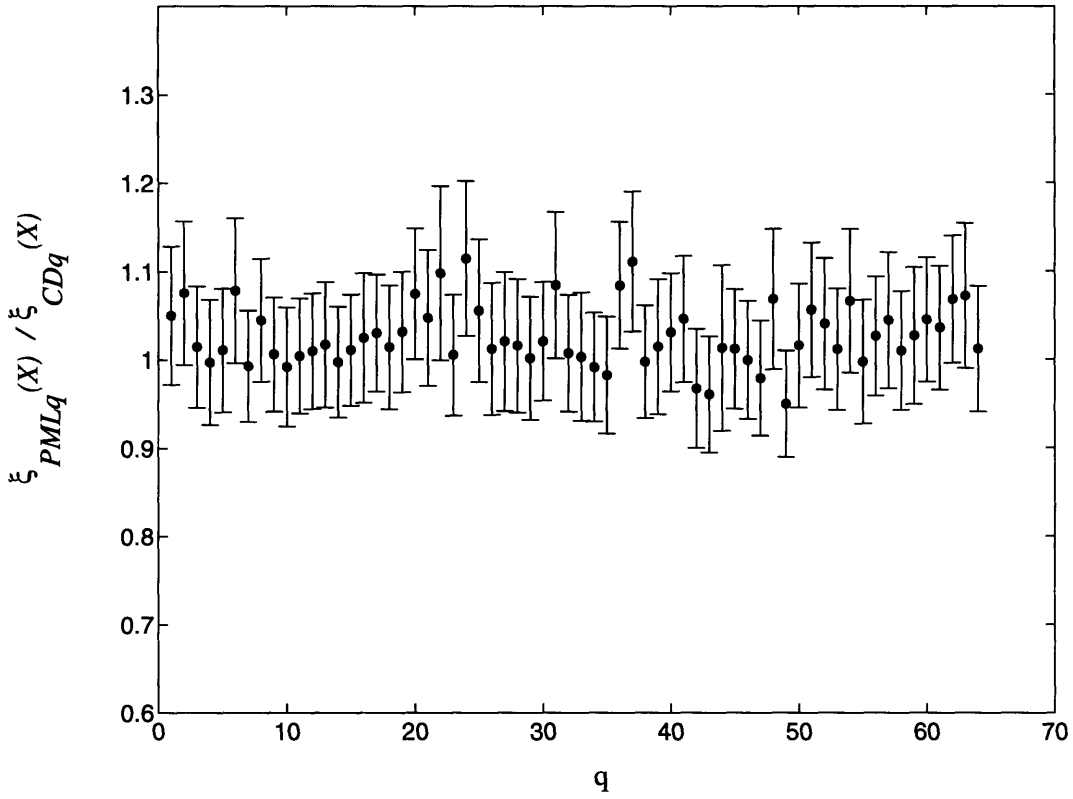

Figure 4-25: Normalized RMS errors of the EM parameter estimates at $P = 64$

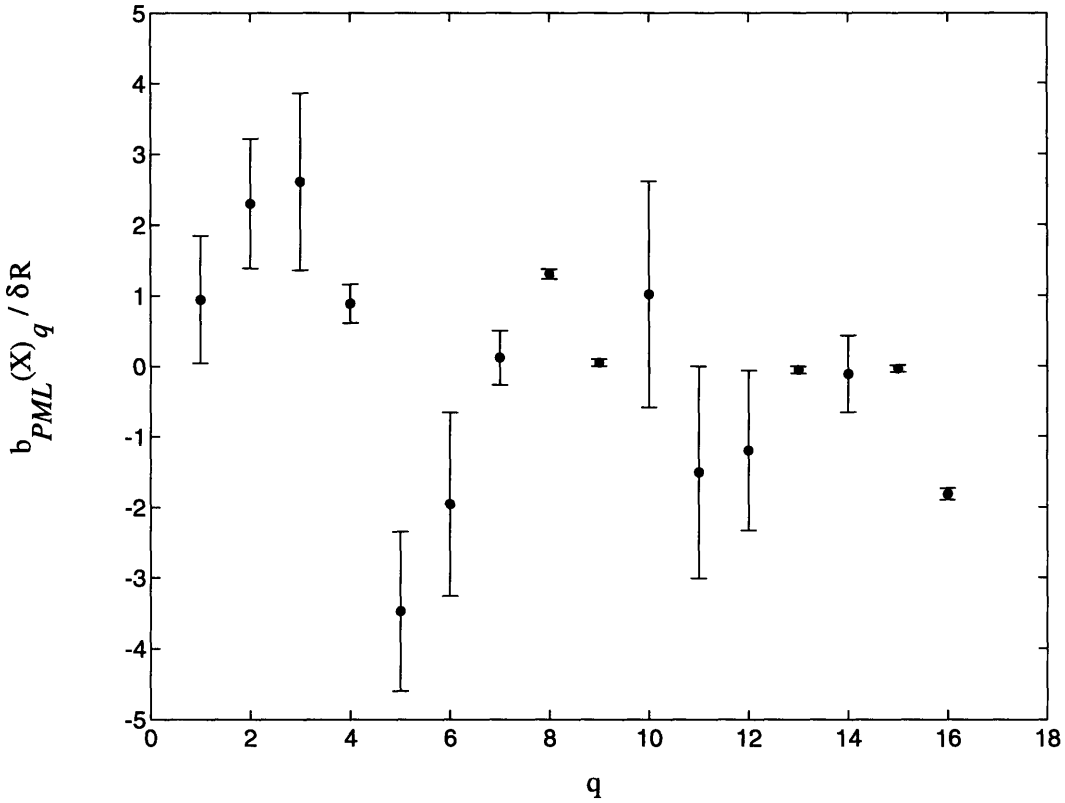

Figure 4-26: Normalized bias of the REM parameter estimates at $P = 16$

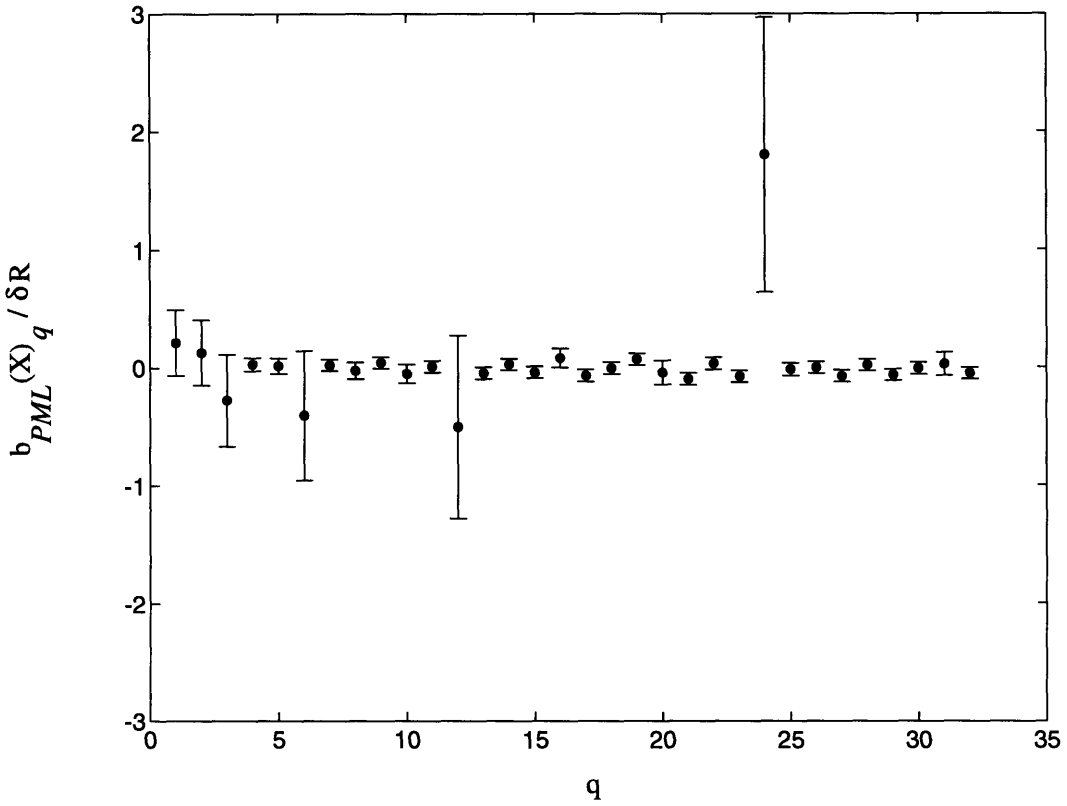

Figure 4-27: Normalized bias of the REM parameter estimates at $P = 32$

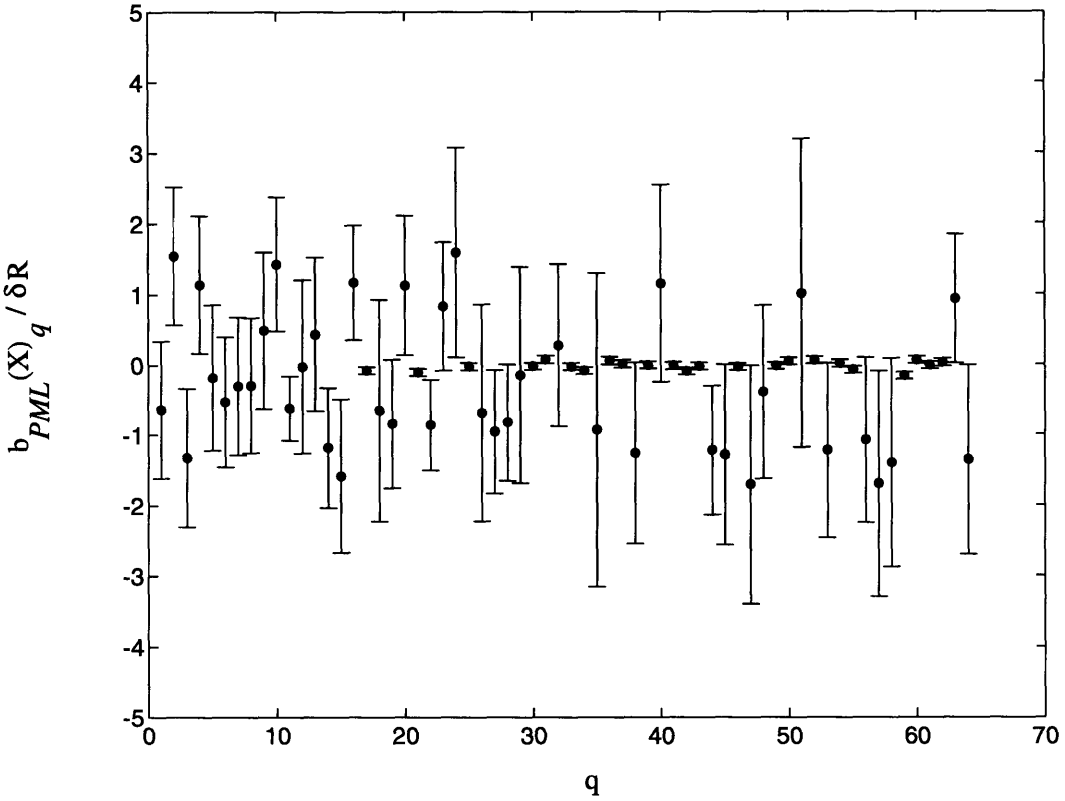

Figure 4-28: Normalized bias of the REM parameter estimates at $P = 64$

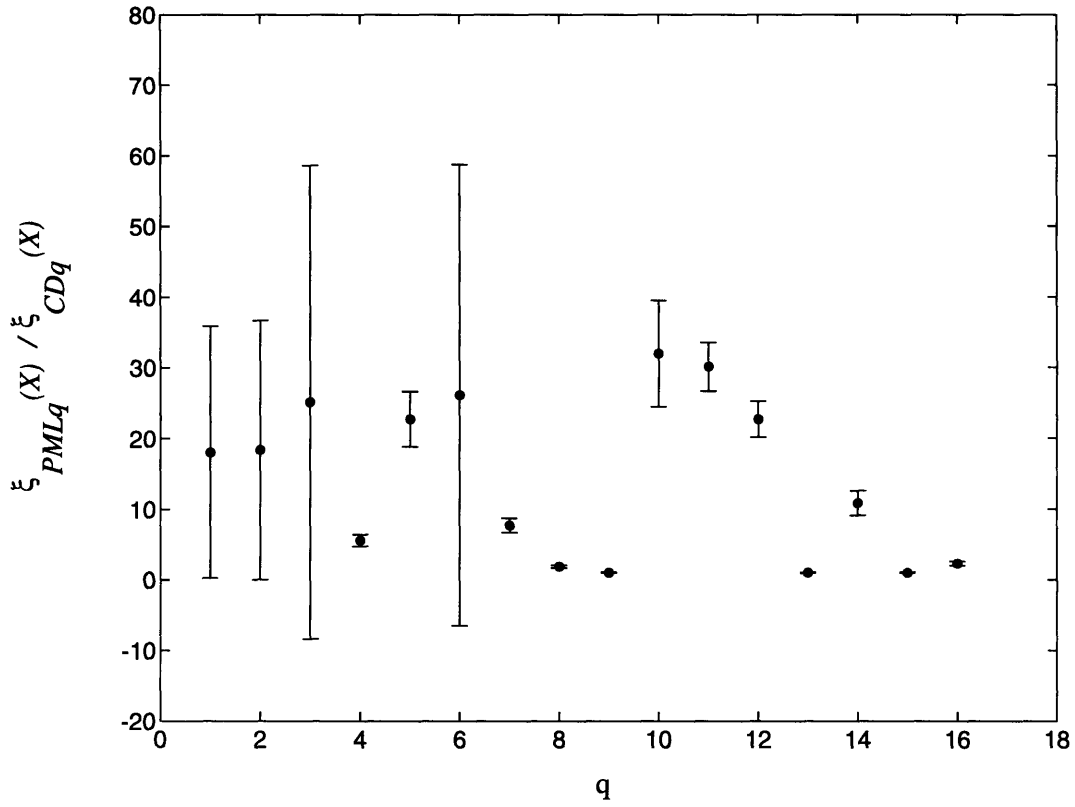

Figure 4-29: Normalized RMS errors of the REM parameter estimates at $P = 16$

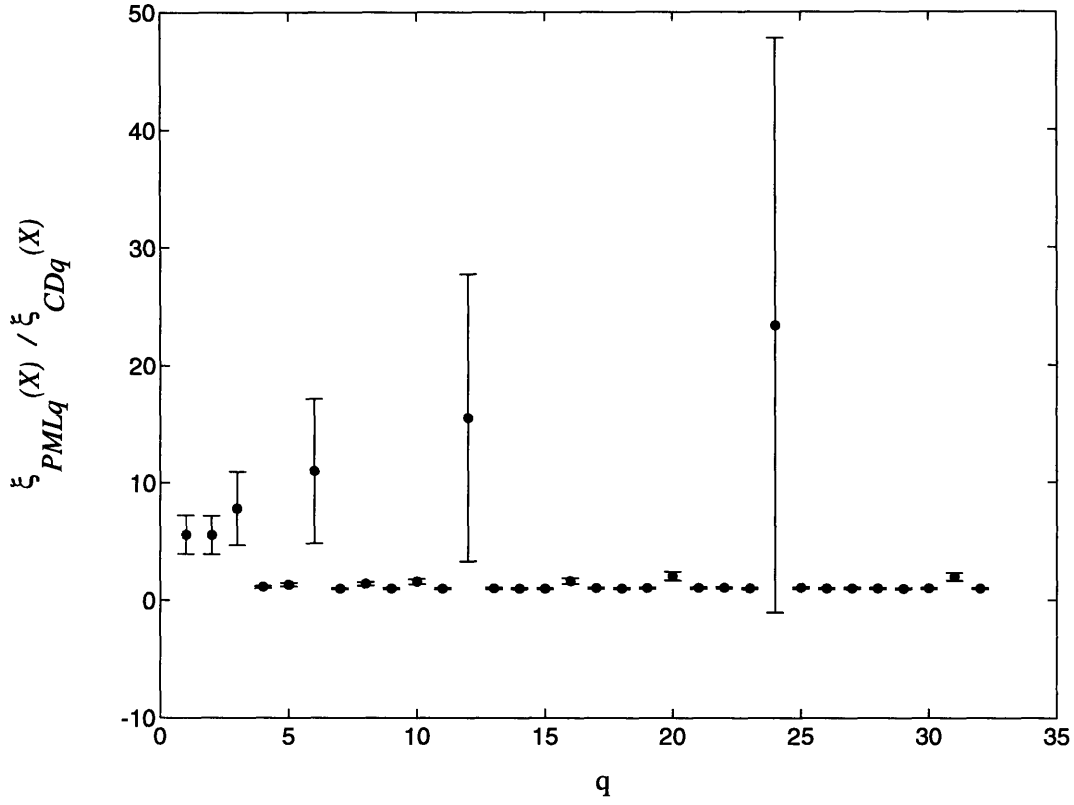

Figure 4-30: Normalized RMS errors of the REM parameter estimates at $P = 32$

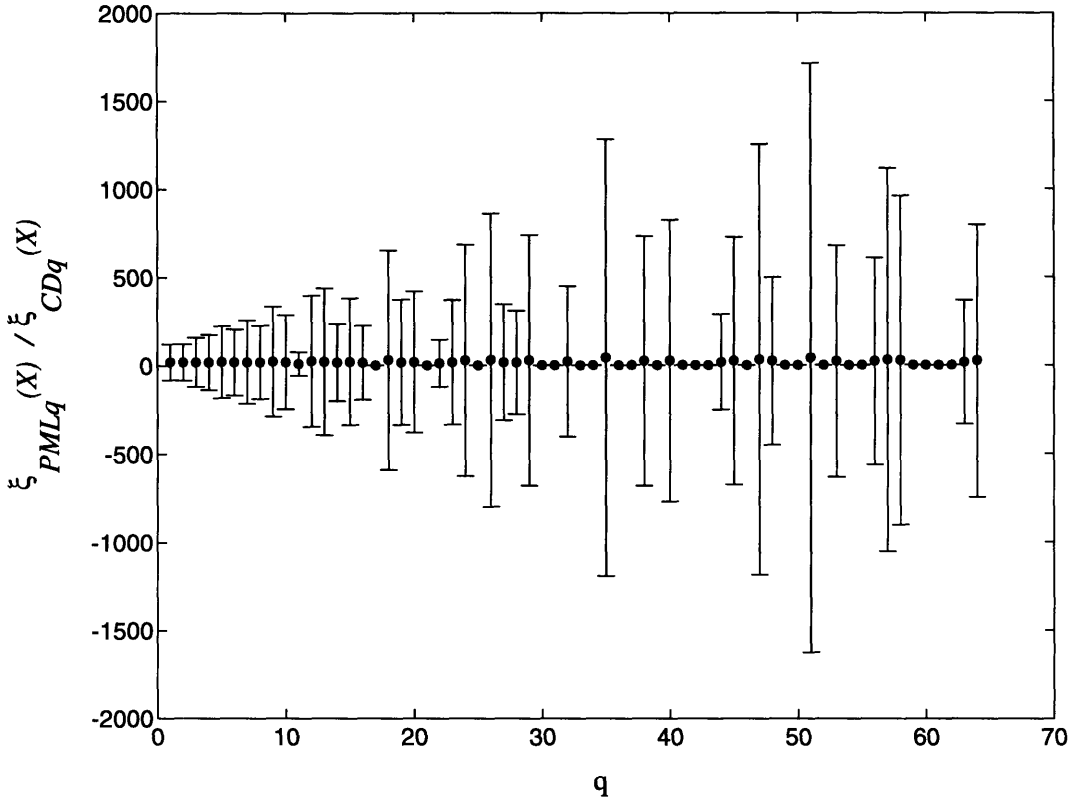

Figure 4-31: Normalized RMS errors of the REM parameter estimates at $P = 64$

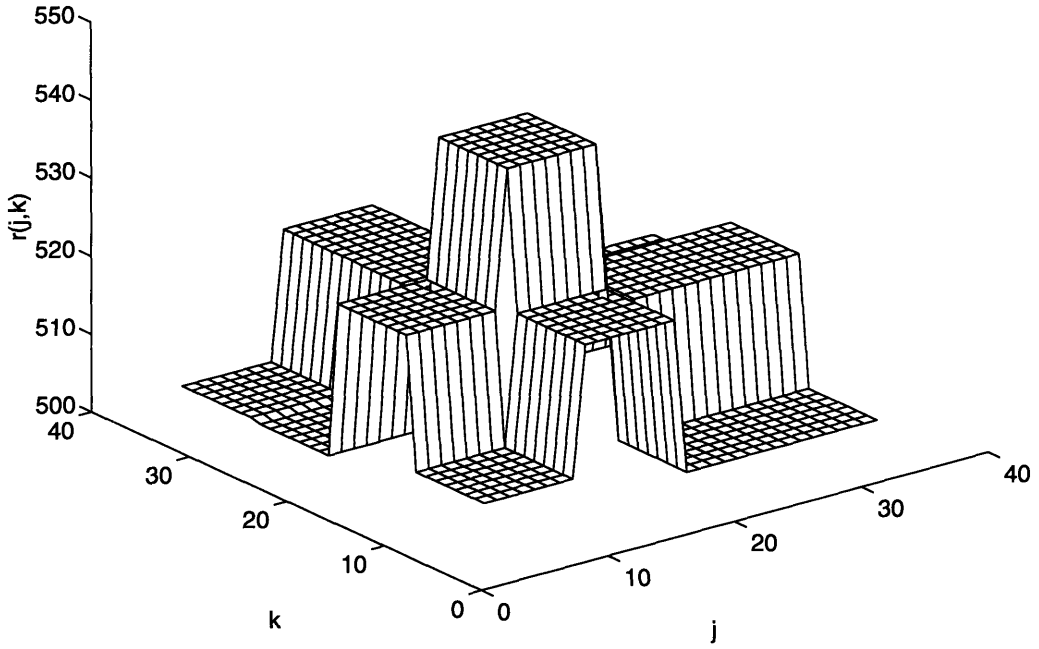

Figure 4-32: 2-D estimated profile with $P_j = 4$, $P_k = 4$ of a 32×32 range image and $\Pr(A) = 0.2$

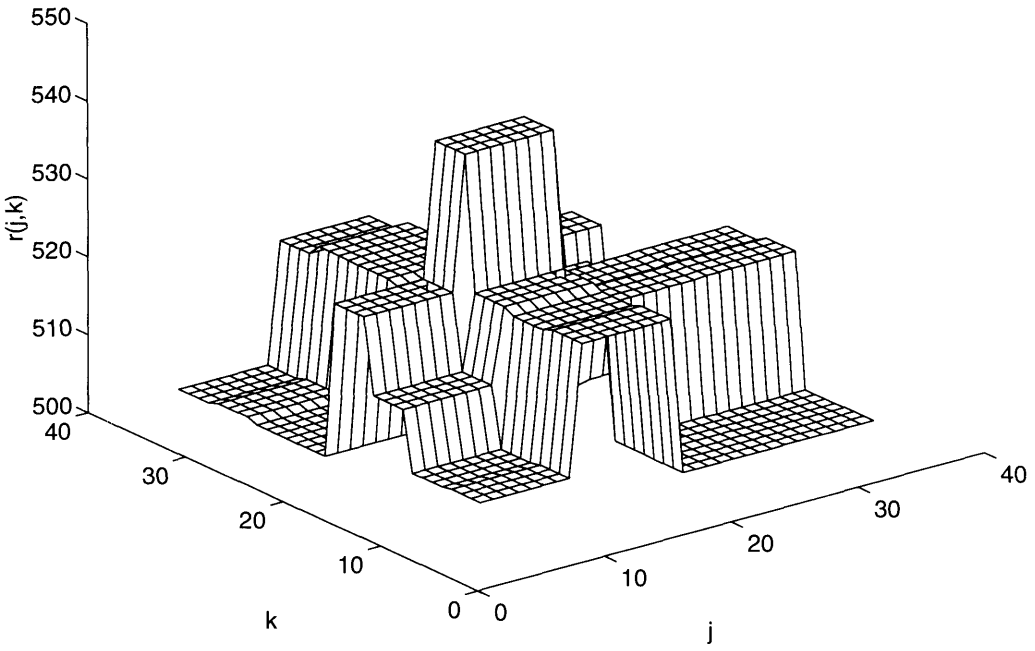

Figure 4-33: 2-D estimated profile with $P_j = 4$, $P_k = 8$ of a 32×32 range image and $\Pr(A) = 0.2$

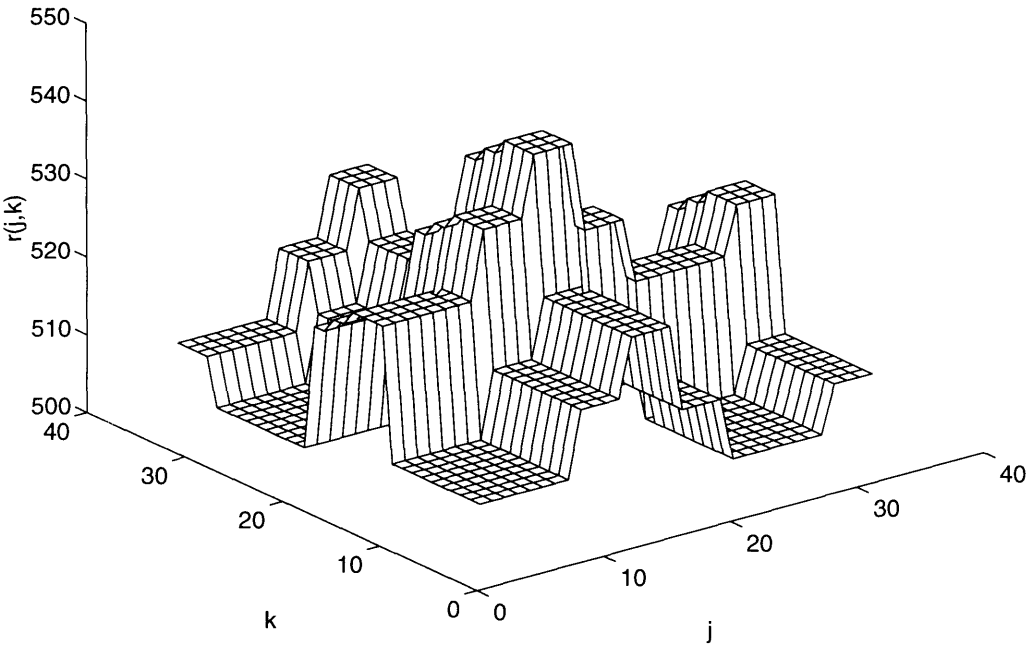

Figure 4-34: 2-D estimated profile with $P_j = 8$, $P_k = 16$ of a 32×32 range image and $\Pr(A) = 0.2$

Chapter 5

Process Termination

The multiresolution, wavelet-based ML profiling scheme allows us to sequence through a progression of estimates of increasingly-fine resolutions. The associated likelihoods also increase with increasing resolution, achieving their maximum at $P = Q$. This means that the full-resolution estimate is the best possible ML estimate, if likelihood is the only performance criterion. However, as noted at the outset of this thesis, anomaly suppression is a critical component of range-image processing. Thus, there is going to be an inevitable trade-off between increasing the resolution (to capture finer scale range-truth details in our ML estimate) and minimizing resolution (to better suppress anomalous pixels). This leads to the question of determining at what scale should this progression terminate to yield an estimate with sufficient dimensionality and satisfactory anomaly suppression.

In the expectation step of the EM algoritm, a weight is computed for each pixel, which determines whether the pixel should be discarded when updating the next estimate. In other words, these weights are indicators of whether of not the corresponding pixel is anomalous, thus providing us a way to measure the quality of the estimate at that particular resolution and $\Pr(A)$. It also suggests a method of terminating this coarse-to-fine ML estimation process. We will begin by studying the statistics of the weights.

5.1 Zero-Weight Statistics

Consider the anomaly data vector given in Eqs. 3.28 – 3.29. Each data point can be interpreted as an independent, identically-distributed *Bernoulli* random variable with a probability of success of $1 - \Pr(A)$ [19]. Suppose we let N_a be the sum of anomalies that occur in the corresponding Q-D range-data vector. It is not difficult to show that the probability mass function (PMF) of this random variable N_a is the binomial, viz.,

$$\Pr(N_a = n) = \binom{Q}{n} \Pr(A)^n (1 - \Pr(A))^{Q-n}, \qquad n = 0, 1, \cdots, Q, \tag{5.1}$$

where

$$\binom{Q}{n} = \frac{Q!}{(Q-n)!n!}. \tag{5.2}$$

From this PMF, we can easily compute the mean,

$$E(N_a) = Q \Pr(A), \tag{5.3}$$

and the standard deviation,

$$\sigma_{N_a} = \sqrt{Q \Pr(A)(1 - \Pr(A))}, \tag{5.4}$$

of N_a. The anomaly data is actually the missing part of the complete data vector and is not available to us. However, the weights in the EM iterations provides us an indirect estimate of the anomalies present.

At each expectation step of the EM algorithm, a weight matrix, $\mathbf{W}_P(\mathbf{X}_P) \equiv \text{diag}[w_q(\mathbf{X}_P)]$, is produced using Eqs. 3.23-3.24. Each of these weights, w_q, is the conditional probability that the associated pixel is not anomalous given that the parameter vector is $\hat{\mathbf{x}}_{PML}$. In other words, the smaller w_q is, the more likely the pixel is anomalous and is thus discarded. In general, all these weights are either very close to zero or one. This behavior is illustrated in Figs. 5-1 – 5-6, which show the weight distributions associated with the EM and REM estimates from Figs. 4-9 – 4-11 and

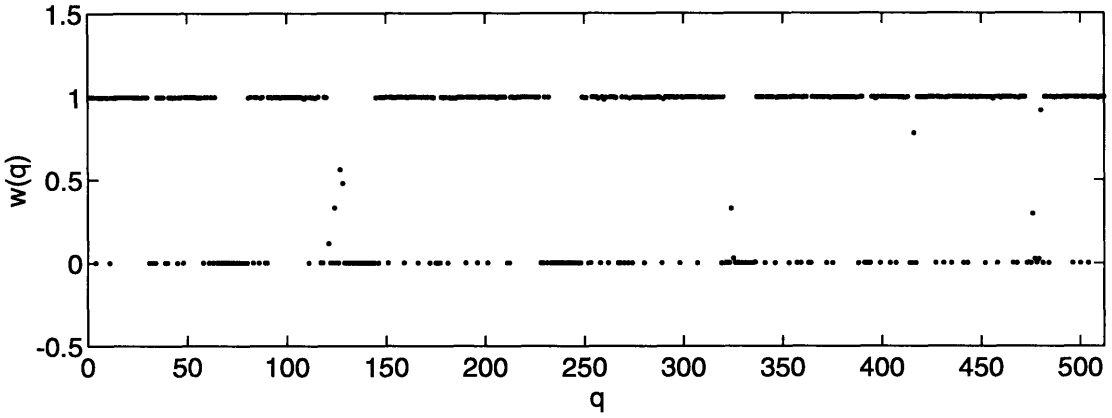

Figure 5-1: Weight distribution w_q vs. q for an EM estimated profile at $P = 16$.

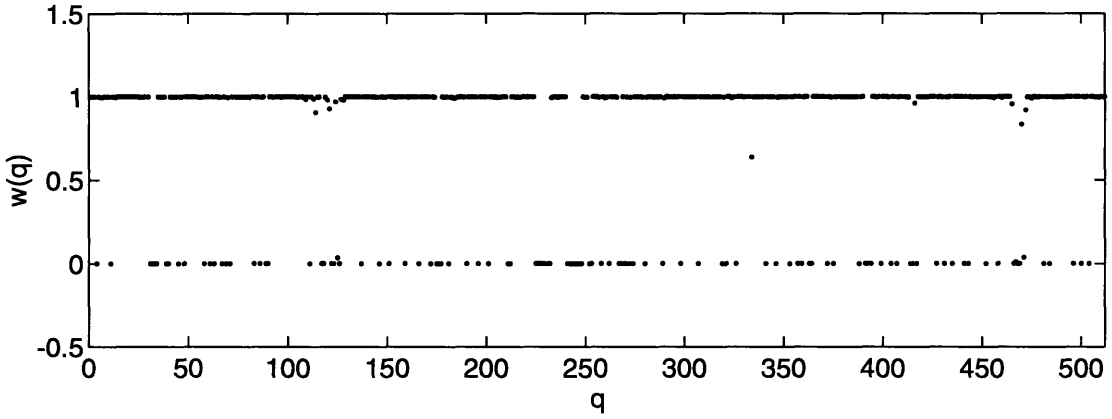

Figure 5-2: Weight distribution w_q vs. q for an EM estimated profile at $P = 32$.

Figs. 4-17 – 4-19, respectively.

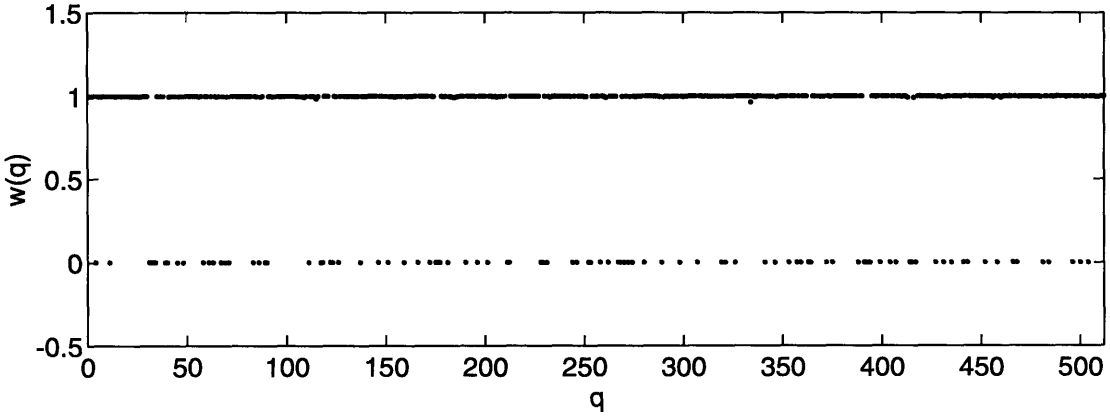

Figure 5-3: Weight distribution w_q vs. q for an EM estimated profile at $P = 64$.

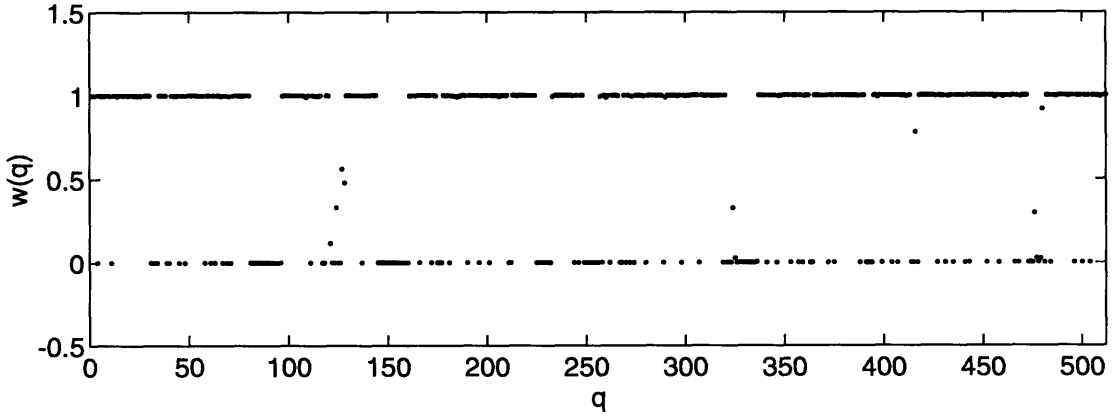

Figure 5-4: Weight distribution w_q vs. q for an REM estimated profile at $P = 16$.

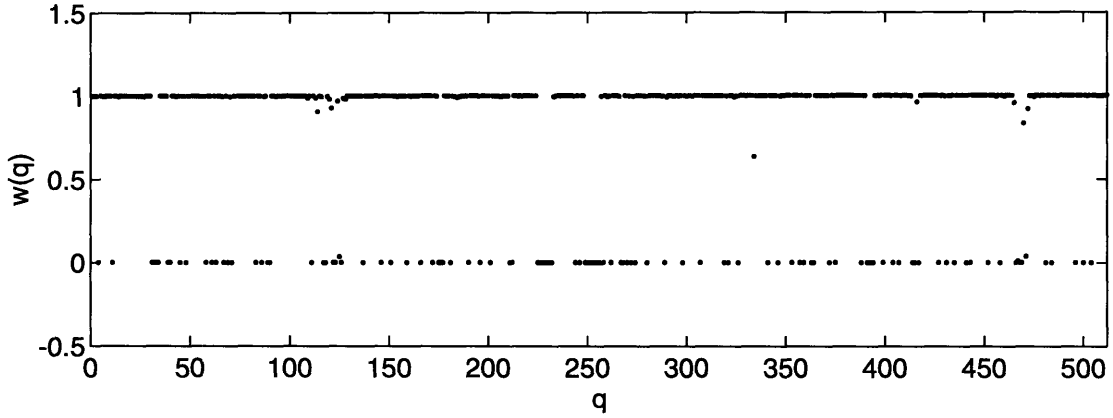

Figure 5-5: Weight distribution w_q vs. q for an REM estimated profile at $P = 32$.

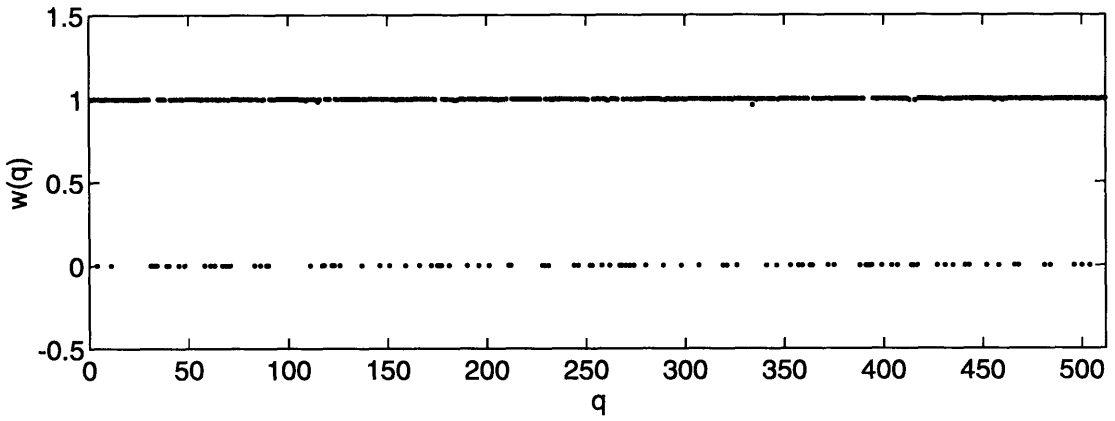

Figure 5-6: Weight distribution w_q vs. q for an REM estimated profile at $P = 64$.

Suppose we binarize the weights by a simple threshold test, i.e., we define

$$z_q = \begin{cases} 0 & \text{if } w_q \leq 0.5, \\ 1 & \text{if } w_q > 0.5, \end{cases} \tag{5.5}$$

which is equivalent to saying

$$z_q = \begin{cases} 0 & \text{if pixel is presumed anomalous,} \\ 1 & \text{if pixel is presumed not anomalous.} \end{cases} \tag{5.6}$$

Thus we can see that $\{z_q; 1 \leq q \leq Q\}$ is also a series of Bernoulli trials. In analogy with N_a, we define N_z to be the number of z_q that are zero, for $1 \leq q \leq Q$. At the appropriate scale P, where we want to terminate the estimation process, we expect to be rejecting approximately the correct number of anomalous pixels by setting their weights to zero. Thus the distribution of N_z would be very similar to that of N_a,

$$\Pr(N_z = n) = \binom{Q}{n} \Pr(A)^n (1 - \Pr(A))^{Q-n}, \qquad n = 0, 1, \cdots, Q. \tag{5.7}$$

Based on the central limit theorem, we can approximate the binomial distribution of N_z with a Gaussian distribution with mean $Q\Pr(A)$ and variance $Q\Pr(A)(1-\Pr(A))$ for $0 \ll Q\Pr(A)(1-\Pr(A)) \ll Q$[19].

With the multiresolution approach, we expect N_z to be a decreasing function of P, since we are rejecting fewer pixels as resolution increases. Eventually, N_z will be

zero when the full-scale is reached since we already know that the best ML estimate is the raw data in which none of the pixels are presumed anomalous.

5.2 Termination Criterion

Based on the value of the mean and standard deviation of the zero-weights, a terminating criterion can be established. The estimation process begins with the coarsest scale and proceeds down to increasingly finer scales. The number of zero-weights of the weight matrix is computed after the last iteration of the EM algorithm at each resolution. Depending how close this value is compared with the theoretical one obtained in Eq. 5.3, we can decide whether we should stop or go to a finer resolution.

Figs. 5-7 – 5-9 show the zero-weight distributions at different resolutions for 500 trials of our the 1-D range profiles. We see that the distributions of N_z are approximately Gaussian in all cases. Both the mean and the standard deviation approach their theoretical values of $Q\,\Pr(A)$ and $\sqrt{Q\,\Pr(A)(1-\Pr(A))}$, respectively. Similar performances are obtained using the REM algorithm, as shown in Figs. 5-10 – 5-12. The mean values calculated are similar to those of the range-truth initialized ones.

The above results suggest a measure for our termination criterion. Assuming we know $\Pr(A)$, the theoretical mean and standard deviation of N_a are available. Note that both values are independent of the resolution P. The value of N_z is evaluated at each resolution as we progress along the coarse-to-fine estimation process, terminating at the coarsest scale where N_z is within one standard deviation of $E(N_a)$. The termination resolution for the set of range-data profiles in our simulations is at $P = 64$, agreeing with the number of parameters in the range truth.

Similar performances are obtained using the REM algorithm, as shown in Figs. 5-10 – 5-12. The mean values calculated are similar to those of the range-truth initialized ones.

The behavior of the weights in our 2-D simulations are slightly different. In general, the zero-weights exhibit similar statistical behavior. The mean and the standard deviation approach the theoretical values as the resolution increases. Thus the same

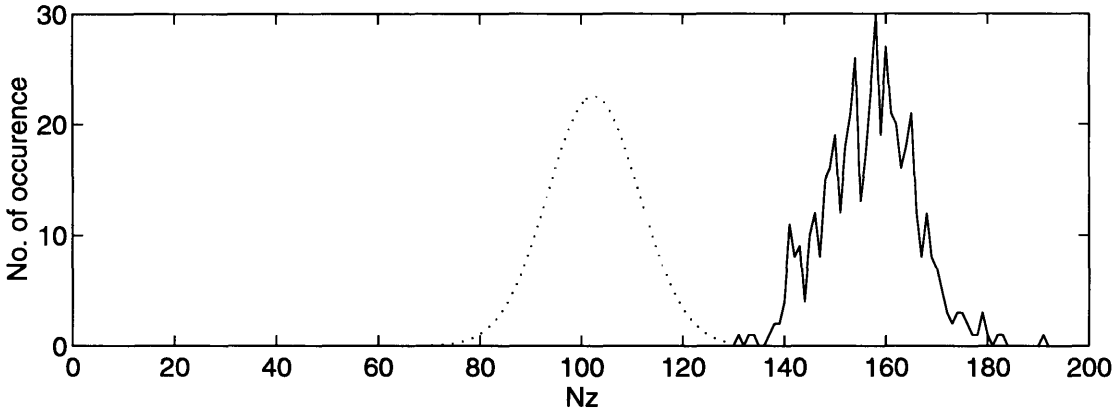

Figure 5-7: Zero-weights distribution with number of occurrences of N_z in 500 trials vs. N_z for an EM estimated profile at $P = 16$ where $E(N_z) = 156.86$ and $\sigma_{N_z} = 8.90$. The dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

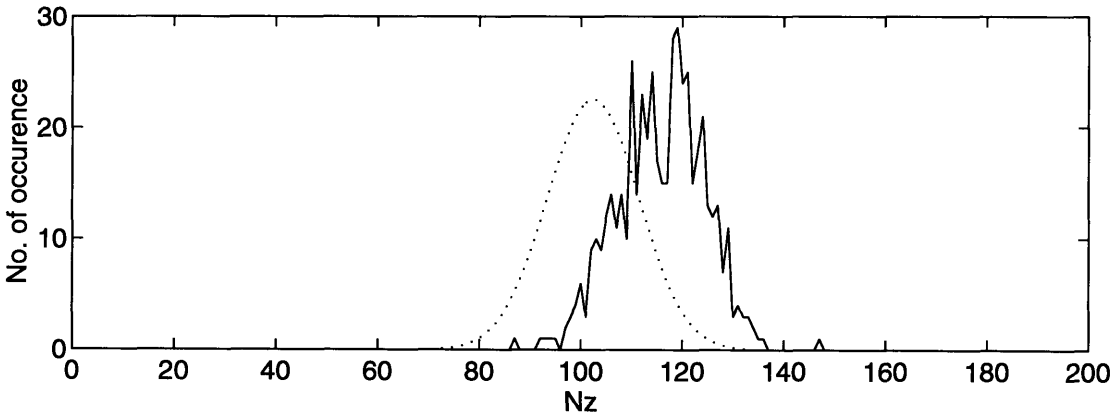

Figure 5-8: Zero-weights distribution with number. of occurrences of N_z in 500 trials vs. N_z for an EM estimated profile at $P = 16$ where $E(N_z) = 115.80$ and $\sigma_{N_z} = 8.49$. The dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

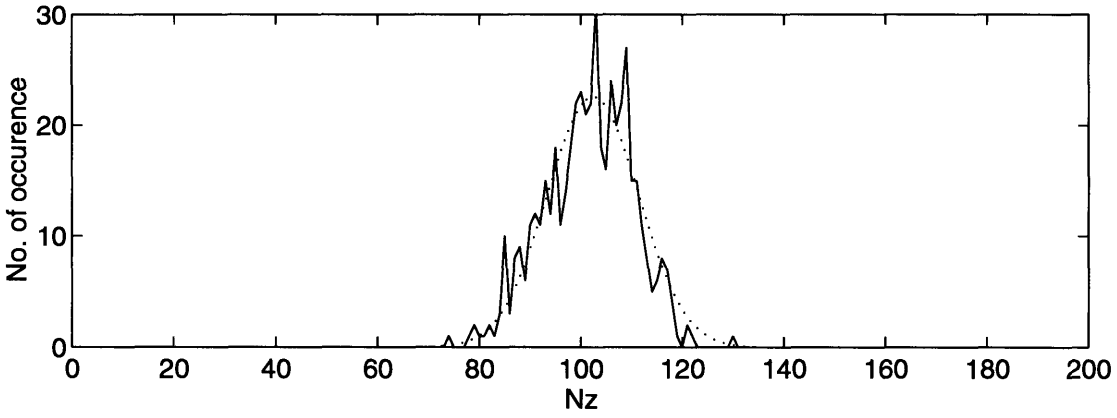

Figure 5-9: Zero-weights distribution with number of occurrences of N_z in 500 trials vs. N_z for an EM estimated profile at $P = 16$ where $E(N_z) = 101.48$ and $\sigma_{N_z} = 8.66$. The dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

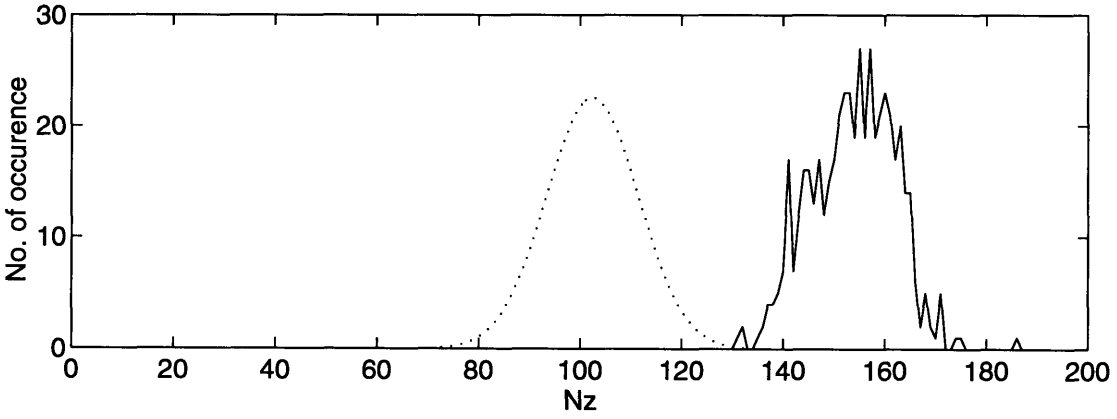

Figure 5-10: Zero-weights distribution with number of occurrences of N_z in 500 trials vs. N_z for an EM estimated profile at $P = 16$ where $E(N_z) = 153.71$ and $\sigma_{N_z} = 8.18$. The dash-dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

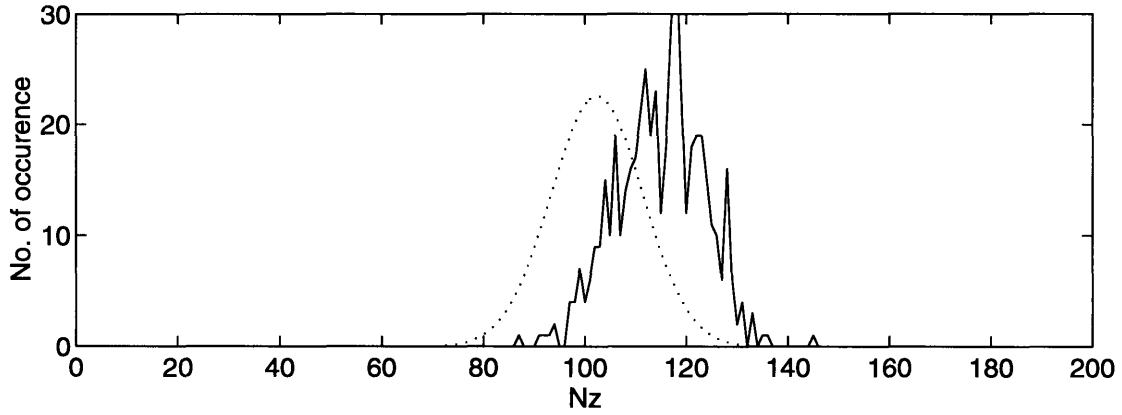

Figure 5-11: Zero-weights distribution with number of occurrences of N_z in 500 trials vs. N_z for an EM estimated profile at $P = 16$ where $E(N_z) = 114.78$ and $\sigma_{N_z} = 8.50$. The dash-dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

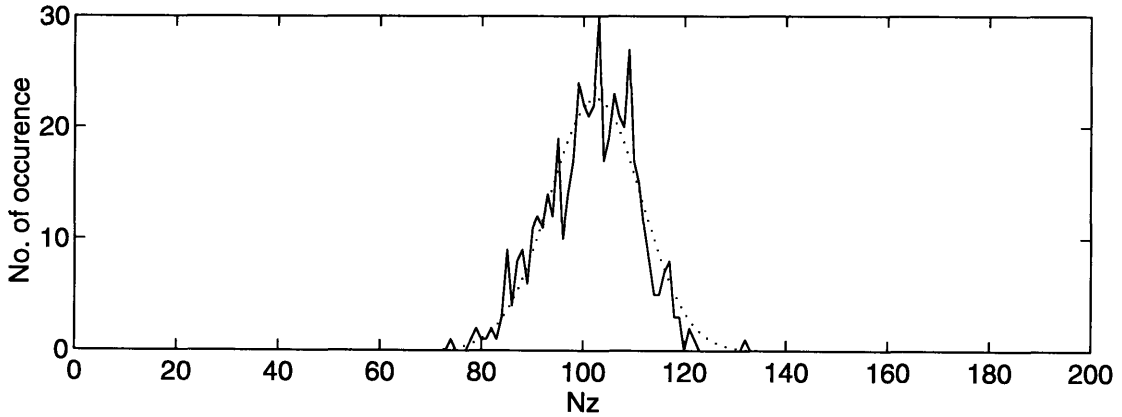

Figure 5-12: Zero-weights distribution with number of occurrences of N_z in 500 trials vs. N_Z for an EM estimated profile at $P = 16$ where $E(N_z) = 101.53$ and $\sigma_{N_z} = 8.70$. The dash-dotted line is the theoretical Gaussian distribution for N_a, with $\Pr(A) = 0.2$, mean $E(N_a) = 102.4$ and standard deviation $\sigma_{N_a} = 9.05$.

termination criterion should be applicable to the 2-D profiles. However, the decision-making is not as straigthtforward as in the 1-D case owing for the following two reasons.

The resolutions of the two 1-D wavelet basese used for the construction of the 2-D bases are independent. Using the notation from Sec. 4.1.2, a given $P = P_j P_k$ can be achieved with a variety of P_j and P_k values. This adds another degree of freedom to the choice of the optimal scale and thus complicates the problem of terminating the estimation process. A few combinations of P_j and P_k may be necessary to determine the appropriate resolution.

The Haar-wavelet basis used in the 2D simulations is a separable basis constructed from two 1D bases. Thus it is best-suited for profiles which have rectangular block features that are aligned with the two axes. In that case, the stopping criterion described above would have the best performance. However, for features of other shapes and alignments, the Haar basis is not intrinsically a good representation and therefore we should not be surprised to see the stopping criterion fail. An example of the worst case would be a diagonal ridge running across the range image. To accurately represent the ridge with the Haar-wavelets, we need to over-resolve the ridge itself, because of its diagonal orientation. In that case, the weight-determined stopping method may not be effective.

In general, the weights generated in the EM algorithm provide a reasonably direct method of determining the quality of the estimate at a particular resolution. The weight-based termination method is applicable to all wavelet bases. However, more sophisticated termination methods can be established for other wavelet-bases and image scenarios.

Chapter 6

Smoothing and Refinement

The ultimate goal of the multiresolution ML estimation process is to obtain a wavelet-fit profile of minimal dimensionality comprising the maximum resolution capability implied by the original image's single-pixel statistics. The final estimate obtained using the termination criterion described in the previous section should have a sufficiently high likelihood and anomaly rejection. Due to the restriction imposed on P, where $P \equiv 2^p < Q$ for any integer p, small range-profile variations (of the order of δR) that are introduced by shot noise are also captured by the fine-scale parameters. These insignificant details can very well be suppressed in most circumstances without affecting the quality of the estimate. Two such techniques are discussed in this chapter.

We assume that the ML estimation process has been successfully terminated by the weight-based stopping criterion developed in the previous chapter. We now have an ML estimate that includes some extraneous fine-scale parameters that should be suppressed. We begin by looking at a piecewise smoothing method in the following section.

6.1 Piecewise Smoothing

The piecewise smoothing method is a simple, straightforward way of suppressing extraneous fine-scale variations. For the moment, let us consider the 1-D case. Basically,

the estimated profiles is divided into several piecewise linear segments along the sharp discontinuities, or edges as we call them, that occur in the profile. We can use the backward difference of the 1-D profile,

$$\delta\hat{r}_q \equiv \hat{r}_q - \hat{r}_{q-1}, \qquad q = 2, 3, \cdots, Q. \tag{6.1}$$

to locate the I points on the estimated profile where $|\delta\hat{r}_q| > t$ for $1 \leq q \leq Q$ and t is around one or two times the local range accuracy. Let the set of these edgepoints be $\{d_i : 0 \leq i \leq I\}$ where $d_0 = 1$ and $d_1 < d_2 < \cdots < d_I$. Upon locating all the edges, the original values of the pixels in each segment are replaced by the respective local means,

$$\hat{r}_s(q) = \frac{1}{d_{i+1} - d_i + 1} \sum_{j=d_i}^{d_{i+1}} r_j, \qquad d_i \leq q \leq d_{i+1}. \tag{6.2}$$

This resulting profile is again fitted with the wavelet basis.

$$\hat{\mathbf{x}}_{P_s} = \mathbf{H}_P^T \hat{\mathbf{r}}_s. \tag{6.3}$$

Parameters that represent the extraneous detail variations are forced to zero at this point, leaving behind the more significant ones that represent the major features in the original range profile. For 2-D profiles, a similar method can be employed. In 2-D, a pixel is grown into a rectangular region, where extraneous fine-scale variations are suppressed. In the following, we will illustrate the piecewise-smoothing process in the simpler, linear-region growing of 1-D data.

Fig. 6-1 shows an estimated profile at resolution 64 which is smoothed with this method. The dimension of the resulting nonzero components of the estimated $\mathbf{X}$ parameter is reduced to 31 by this procedure. Comparing with the true parameter vector, there are only 6 additional nonzero components. In terms of computation and operation complexity, the piecewise smoothing method is very efficient. This is desirable when scale can be traded off for dimensionality.

Figs. 6-2 and 6-3 show the performance of this smoothing algorithm over 500 trials

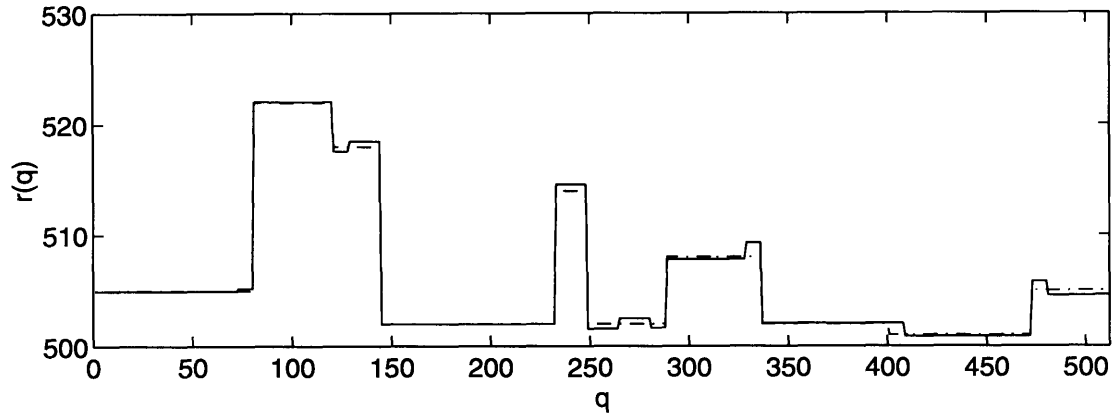

Figure 6-1: Piecewise-smoothed estimated profile at $P = 64$; dash-dotted line is the range-truth.

of range data originating from the range truth shown above. The smoothed estimate actually performs better than the original estimate and the RMS errors go well below the CD bound. Since the CD bound requires that the estimate be unbiased for all $\mathbf{X}$, the result leads us to the conclusion that this "ad-hoc" smoothing technique is biased for the type of profiles with wide features and large discontinuities. This is not hard to understand since the effect of local averaging is more apparent in wide, flat areas with large distinct features. For profiles with slow varying features, this smoothing technique does not have much effect.

Consider an intuitive worst case for our piecewise smoother in which the range-truth resembles a staircase, as shown in Fig. 6-4. All the features are 8 pixels in width and increase in stepsize of $0.75\delta R$. Using the piecewise smoothing method, the smoothed range truth (or a range data without noise) will be a constant profile with a single nonzero parameter. With $\Pr(A) > 0$, certain parts of the smoothed profile will be wider, depending on the distribution of local noise and the anomalies. A typical smoothed profile can be seen in Fig. 6-5 for $\Pr(A) = 0.2$. For this type of profile, the performance of the smoothing method is worse than the original estimate. The bias and the RMS errors of this estimate can be found in Figures 6-6 and 6-7.

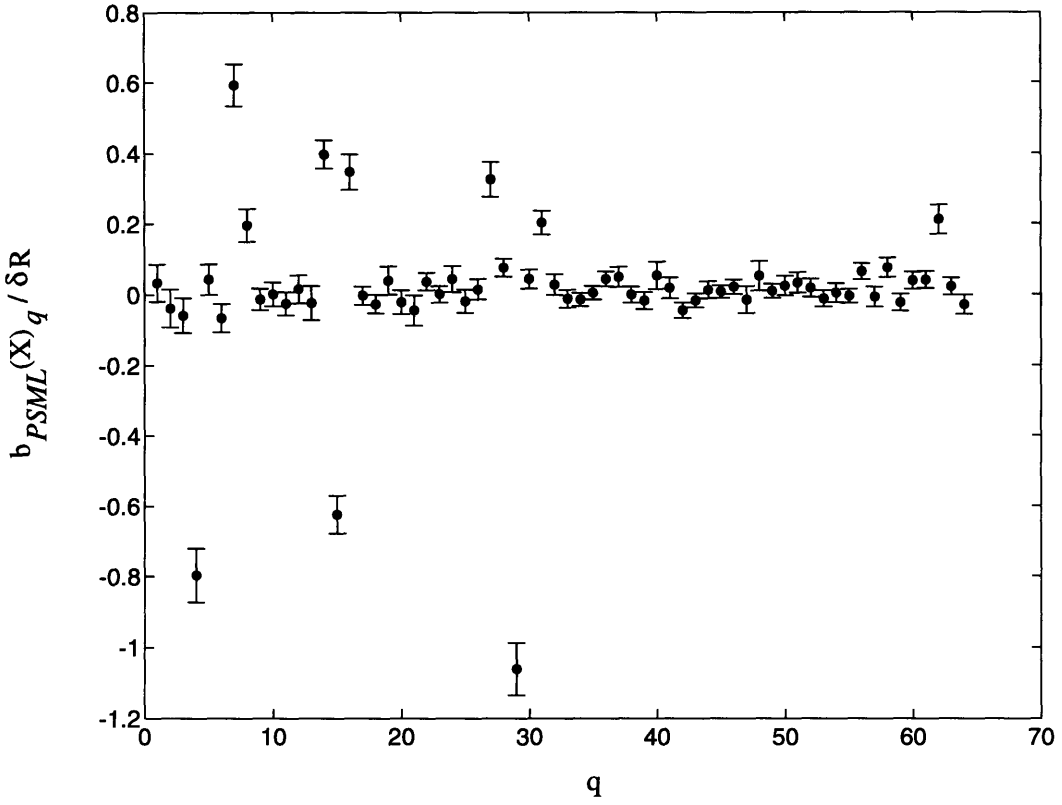

Figure 6-2: Normalized bias of the parameters of the piecewise-smoothed profiles at $P = 64$

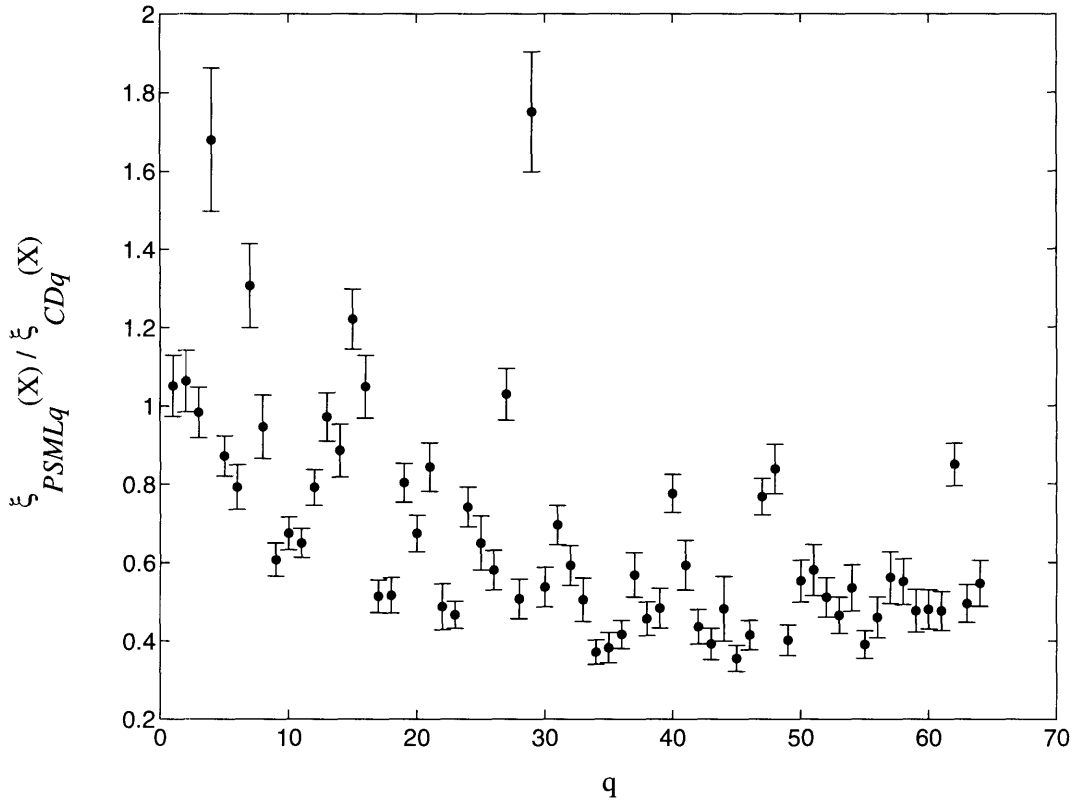

Figure 6-3: Normalized RMS errors of the parameters of the piecewise-smoothed profiles at $P = 64$

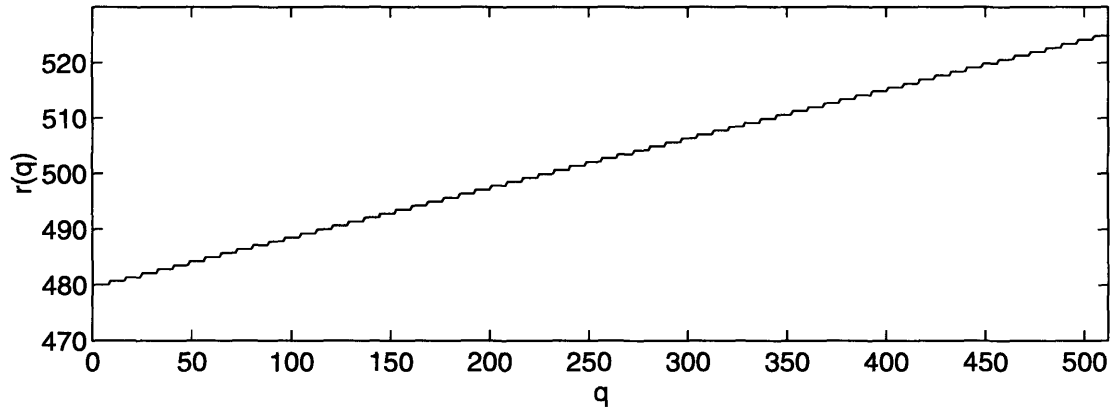

Figure 6-4: A staircase profile with steps 8 pixels wide and $0.75\delta R$ high.

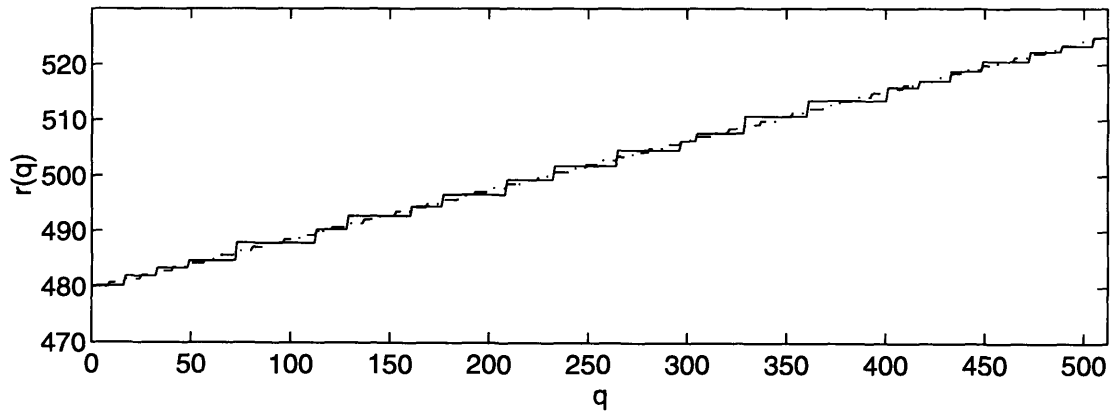

Figure 6-5: Staircase profile estimate after smoothing; dash-dotted line is the range-truth.

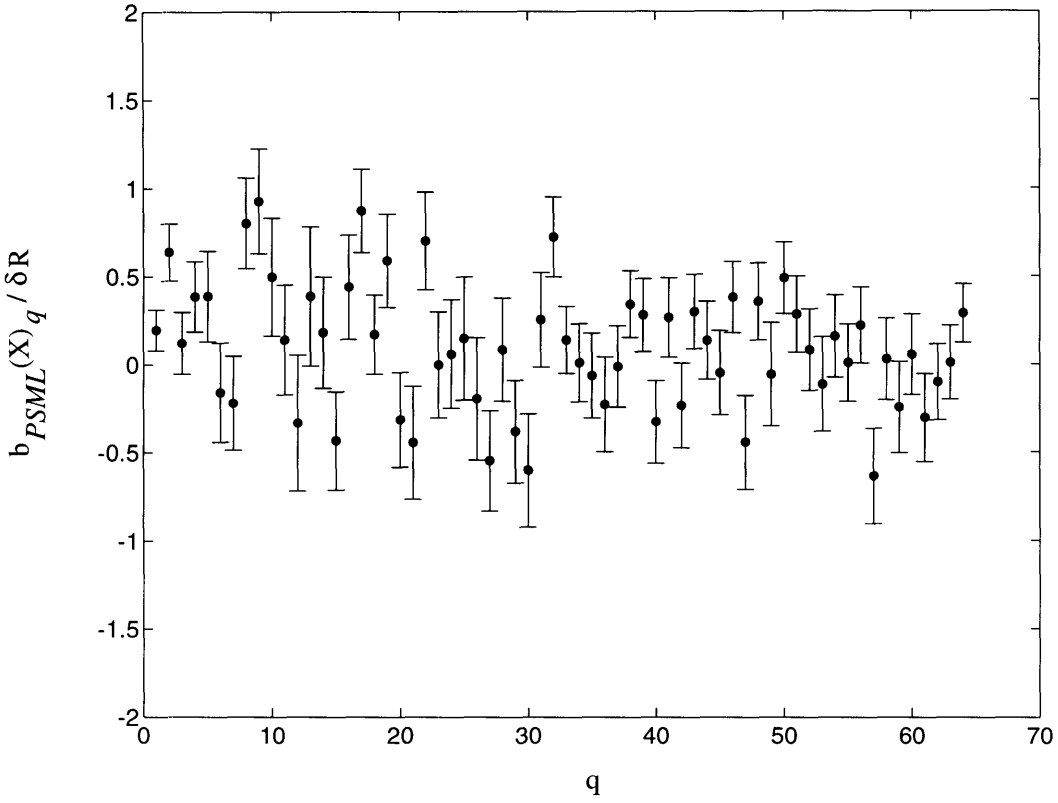

Figure 6-6: Normalized bias of the parameters of the piecewise-smoothed staircase profile at $P = 64$ for 100 trials.

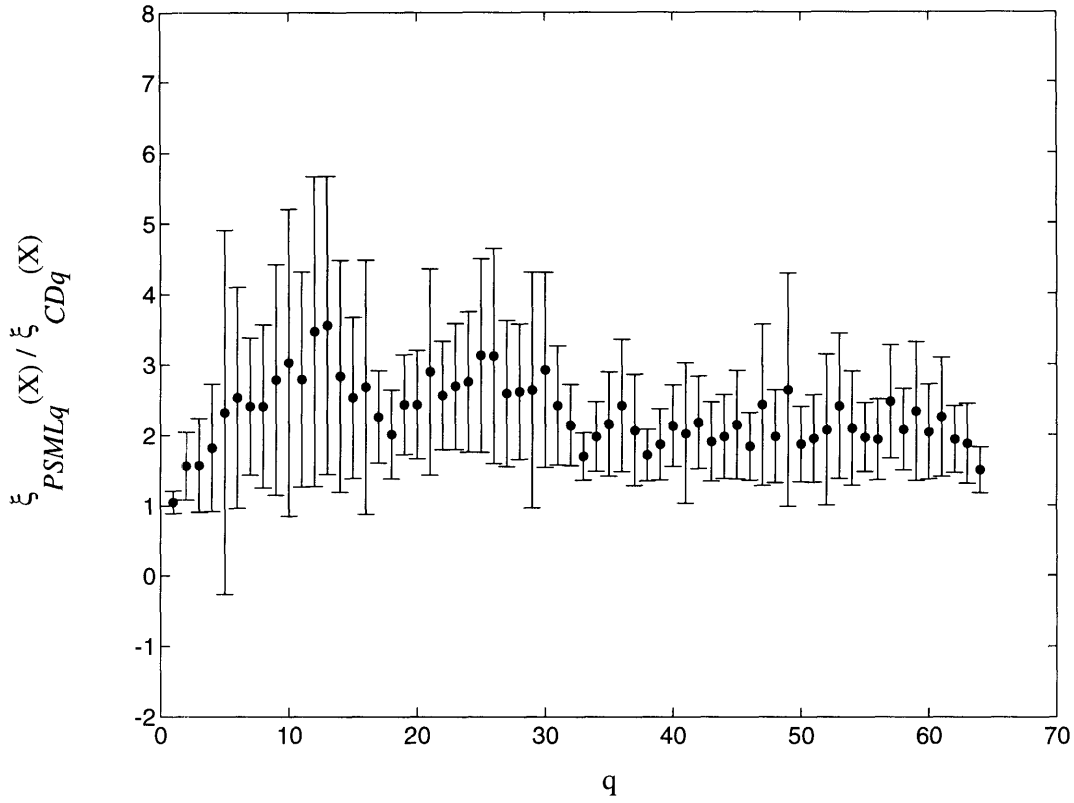

Figure 6-7: Normalized RMS errors of the parameters of the piecewise-smoothed staircase profiles at $P = 64$ for 100 trials.

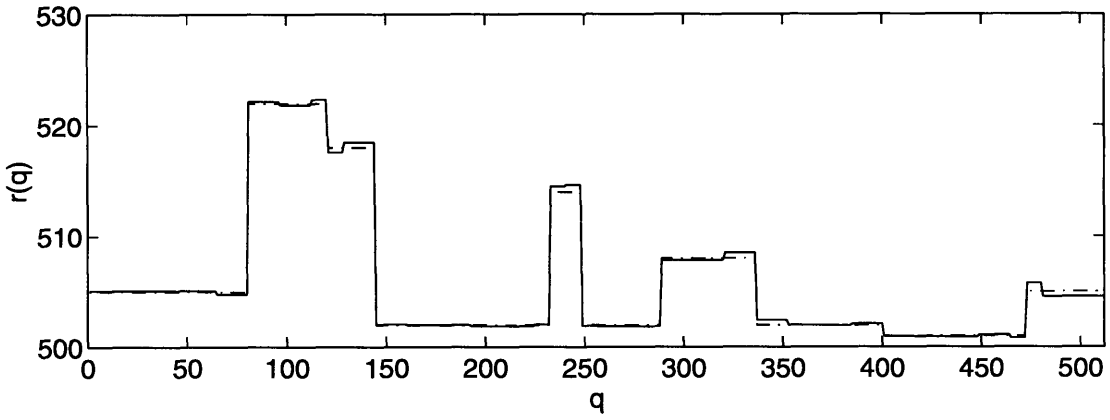

Figure 6-8: Typical profile obtained from estimating only the non-zero parameters of the range truth; dash-dotted line is the range-truth.

6.2 Non-zero Parametric Profiling

In essence, the piecewise smoothing method smoothes the estimated profile by replacing the pixels with the local mean over a segment, thus forcing the smallest fine-scale parameter estimates to zero. In other words, we only want to estimate the parameters that are implied by the true profile. Since the range truth is known in our simulations, it is possible to examine the performance obtainable if we only estimate the nonzero parameters at a specific resolution. In that case, the estimated profile is already at its minimal dimension and no further smoothing process is needed.

Fig. 6-8 shows a ML estimated profile for $\Pr(A) = 0.2$ obtained from the EM algorithm seeded with the range truth. Only the 25 nonzero parameters of the range-truth are estimated. We can see that the resulting profile is already smoother than the one in Fig. 4-11. No additional refinement is necessary since the ML estimation already yields the estimate that has the minimal dimensionality embedded in the range truth.

The performance of this estimation approach is shown in Figures 6-9 and 6-10. The final estimate is both unbiased and efficient, as indicated by the plots. In fact, it even performs better than the ML estimate at our terminal resolution. This is the optimal ML estimate we can get with good anomaly suppression.

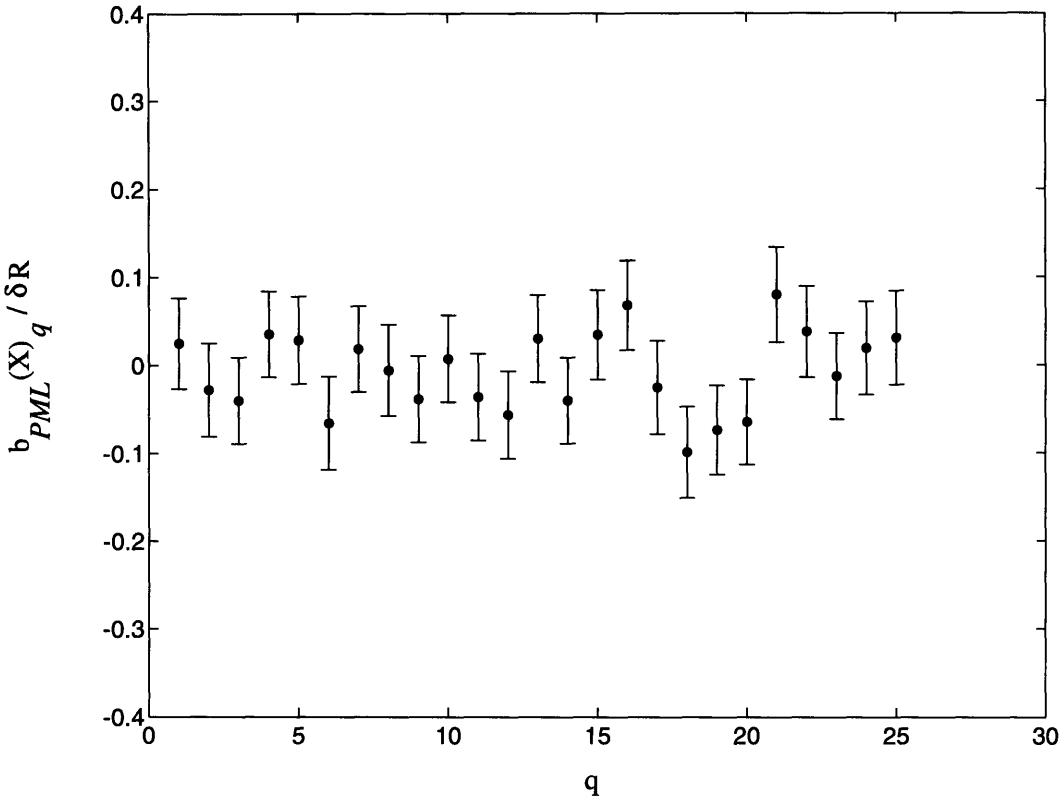

Figure 6-9: Normalized bias of 500 trials of the non-zero parameters estimation.

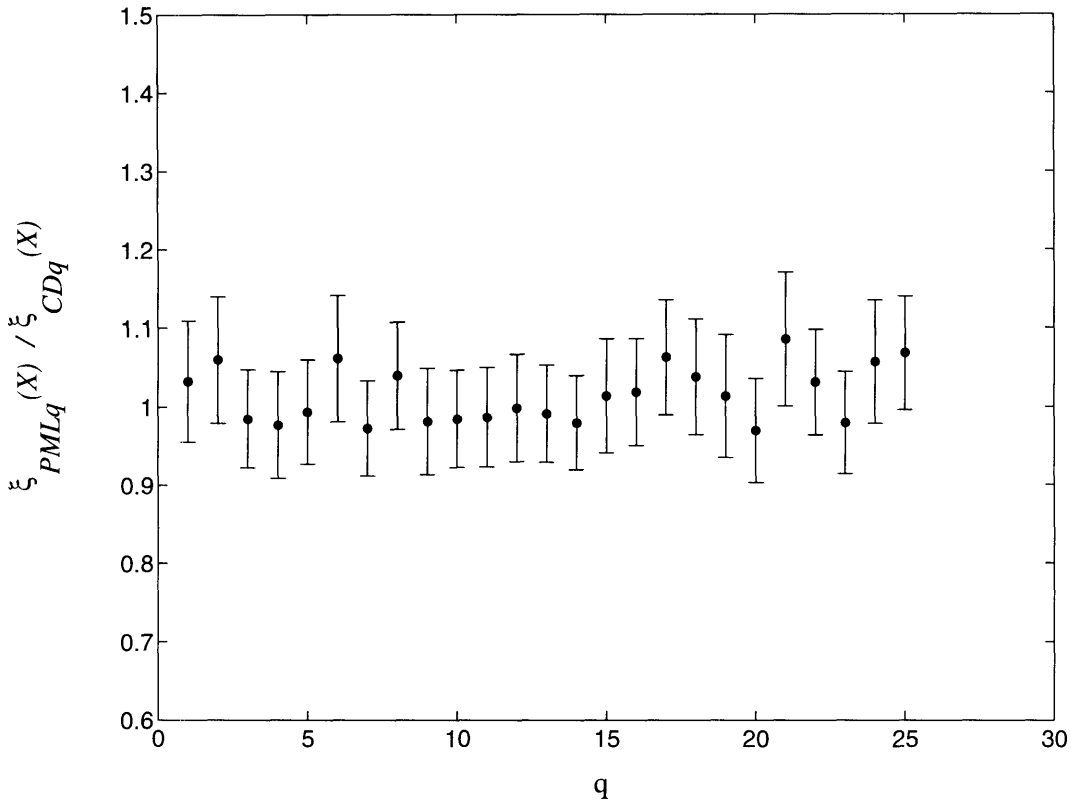

Figure 6-10: Normalized RMS error of 500 trials of the non-zero parameters estimation.

When the true range is not available, as in the case of real laser radar data, the above approach is not feasible as we have no means of knowing which of the parameters are non-zero. However, Figs. 6-8 – fig:nz-rms demonstrate the value of having some prior knowledge of the parameters that we are estimating in reducing the dimensionality of the estimate profile. This leads us logically to the development of a more practical smoothing scheme via maximum a posteriori probability estimation.

6.3 Maximum A Posteriori Probability Estimation

The maximimum a posteriori probability (MAP) estimate $\hat{\mathbf{x}}_{PMAP}$ is the $\mathbf{X}_P$ that maximizes the conditional density $p_{\mathbf{x}_P|\mathbf{r}}(\mathbf{X}'_P \mid \mathbf{R})$:

$$\begin{aligned} \hat{\mathbf{x}}_{PMAP} &= \arg\max_{\mathbf{X}_P} p_{\mathbf{x}_P|\mathbf{r}}(\mathbf{X}_P \mid \mathbf{R}) \\ &= \arg\max_{\mathbf{X}_P} \frac{p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P) p_{\mathbf{x}_P}(\mathbf{X}_P)}{p_{\mathbf{r}}(\mathbf{R})}. \end{aligned} \tag{6.4}$$

Again, it is often easier to maximize the logarithm of the likelihood function. Taking the logarithm and differentiating with respect to $\mathbf{X}_P$, we find that the MAP estimate satisfies

$$\frac{\partial}{\partial X_q} \ln[p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P) p_{\mathbf{x}_P}(\mathbf{X}_P)] \mid_{\mathbf{X}_P = \hat{\mathbf{x}}_{PMAP}} = 0, \qquad 1 \le q \le P. \tag{6.5}$$

This is the same as saying

$$\left[\frac{\partial}{\partial X_q} \ln p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P) + \frac{\partial}{\partial X_q} \ln p_{\mathbf{x}_P}(\mathbf{X}_P) \right] \mid_{\mathbf{X}_P = \hat{\mathbf{x}}_{PMAP}} = 0. \tag{6.6}$$

Suppose we know that $\mathbf{X}_P$ is Gaussian with $N(\mathbf{m}_{\mathbf{x}_P}, \Lambda_{\mathbf{x}_P})$. Following Eqs. 3.22, 3.23 and 6.7, our MAP equation can be reduced to

$$\frac{1}{\delta R^2}\mathbf{H}_P^T\mathbf{W}_P(\mathbf{X}_P)[\mathbf{R}-\mathbf{H}_P\mathbf{X}_P]-(\mathbf{X}_P-\mathbf{m}_{\mathbf{x}_P})^T\Lambda_{\mathbf{x}_P}^{-1}=\mathbf{0}. \tag{6.7}$$

Note that this is very similar to the ML equation we have in Eq. 3.22, except for an additional second term that involves the prior statistics of $\mathbf{X}_P$. Next we need to show that the EM algorithm can be used for the MAP estimation.

Suppose we assume a least-square initial estimate, $\hat{\mathbf{x}}_{PMAP}(0)$. Following the definition of $Q(\mathbf{X'}_P \mid \mathbf{X}_P)$ in [11], let us define the function

$$\begin{aligned}
&Q'(\mathbf{X'}_P \mid \mathbf{X}_P)\\
&\quad\equiv\ Q(\mathbf{X'}_P \mid \mathbf{X}_P)+\ln p_{\mathbf{x}_P}(\mathbf{X'}_P)\\
&\quad\equiv\ \sum_{\mathbf{A}}\Pr(\mathbf{a}=\mathbf{A}\mid\mathbf{r}=\mathbf{R},\mathbf{x}_P=\mathbf{X}_P)\\
&\qquad\times\ln[\Pr(\mathbf{a}=\mathbf{A}\mid\mathbf{r}=\mathbf{R},\mathbf{x}_P=\mathbf{X'}_P)p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R}\mid\mathbf{X'}_P)]+\ln p_{\mathbf{x}_P}(\mathbf{X'}_P),
\end{aligned} \tag{6.8}$$

where the vector $\mathbf{A}$ is summed over all possible anomaly combinations, viz.,

$$\mathbf{A}\equiv\begin{bmatrix}A_1\\A_2\\\vdots\\A_Q\end{bmatrix}. \tag{6.9}$$

Here, $Q'(\mathbf{X'}_P \mid \mathbf{X}_P)$ is essentially the conditional expectation of the complete data log-a-posteriori probability density function evaluated at $\mathbf{x}_P=\mathbf{X'}_P$, given the incomplete data vector is $\mathbf{r}=\mathbf{R}$ and the true parameter vector is $\mathbf{x}_P=\mathbf{X}_P$.

As described in Sec. 3.5, the EM algorithm updates the estimate $\hat{\mathbf{x}}_P(n)$ in two steps. In the expectation step, the conditional expectation $Q'(\mathbf{X'}_P \mid \mathbf{X}_P)$ is computed. With reference to the results in [11], we find that the expectation step can again be done analytically,

$$
\begin{aligned}
Q'(\mathbf{X'}_P \mid \mathbf{X}_P) \;=\; & Q \ln\left[\frac{\Pr(A)}{\Delta R}\right] + \ln\left[\frac{(1-\Pr(A))\Delta R}{\Pr(A)\sqrt{2\pi\delta R^2}}\right] \sum_{q=1}^{Q} w_q(\mathbf{X}_P) \\
& - \sum_{q=1}^{Q} w_q(\mathbf{X}_P) \frac{[R_q - (\mathbf{H}\mathbf{X'_P})_q]^2}{2\delta R^2} - \frac{(\mathbf{X'}_P - \mathbf{m_x})^T \Lambda_{\mathbf{x}} (\mathbf{X'}_P - \mathbf{m_x})}{2} \\
& - \ln[(2\pi)^{P/2} \det \Lambda_{\mathbf{x}}{}^{1/2}]. \qquad (6.10)
\end{aligned}
$$

For the maximization step, we want to set the $\hat{\mathbf{x}}_P(n+1)$ equal to the $\mathbf{X'}_P$ that maximizes $Q'(\mathbf{X'}_P \mid \mathbf{X}_P)$. By differentiating Eq. 6.11, we find that $\hat{\mathbf{x}}_P(n+1)$ must satisfy

$$
\frac{1}{\delta R^2} \mathbf{H}_P^T \mathbf{W}_P(n)[\mathbf{R} - \mathbf{H}_P \hat{\mathbf{x}}_P(n)] - (\hat{\mathbf{x}}_P(n+1) - \mathbf{m}_{\mathbf{x}_P})^T \Lambda_{\mathbf{x}}^{-1} = \mathbf{0}, \qquad (6.11)
$$

where $\mathbf{W}_P(n)$ is the weight matrix calculated based on the previous estimate. Consider the linear Gaussian problem in which we observe

$$
\mathbf{r} = \mathbf{H}_P \mathbf{x'}_P + \mathbf{v} \qquad (6.12)
$$

where $\mathbf{x'}_P \sim N(\mathbf{m_x}, \Lambda_{\mathbf{x}})$ and $\mathbf{v} \sim N(0, \delta R^2 \mathbf{W}_P(n))$. In this case, the Bayes Least Squares Estimate is also the MAP estimate. Specifically, a form of this estimate can be written as [18]

$$
\hat{\mathbf{x}}_{PMAP}(n+1) = \mathbf{m_x} + \Lambda_B(n) \mathbf{H}^T \mathbf{W}_P(n)[\mathbf{R} - \mathbf{H}_P \hat{\mathbf{x}}_{PMAP}(n)], \qquad (6.13)
$$

where

$$
\Lambda_B^{-1}(n) = \Lambda_{\mathbf{x}}^{-1} + \mathbf{H}_P^T \mathbf{W}_P(n) \mathbf{H}_P \qquad \text{for} \quad n = 0, 1, 2, \cdots. \qquad (6.14)
$$

This Bayes LS estimate obeys the same necessary condition given by Eq. 6.11, so it provides the solution to the EM algorithm's MAP maximization step. Comparing with Eq. 3.33, we see that Eq. 6.15 differs in its additional information contributed by the prior knowledge of the statistics of $\mathbf{x}$. Note that as $\Lambda_{\mathbf{x}} \to \infty$, meaning that

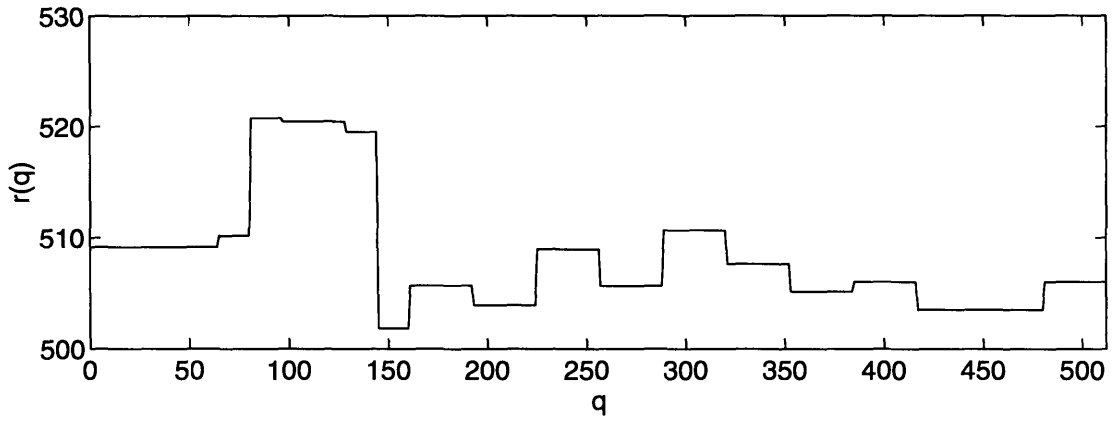

Figure 6-11: Mean range-truth generated by the given $\mathbf{m_x}$ and Λ_x.

there is almost no prior information about $\mathbf{x}_P$, we have

$$\Lambda_B \rightarrow (\mathbf{H}_P^T \mathbf{W}_P(n) \mathbf{H}_P)^{-1}, \tag{6.15}$$

and

$$\hat{\mathbf{x}}_{PMAP}(\mathbf{R}) \rightarrow \hat{\mathbf{x}}_{PML}(\mathbf{R}). \tag{6.16}$$

With some prior knowledge about the relative importance of each element in the parameter-vector $\mathbf{x}_P$, MAP estimation provides us a way to suppress extraneous fine-scale variations and obtain a smoother final estimate. More importantly, the EM algorithm can be used with minimal alterations in the implementation. Figs. 6-13 – 6-15 show the resulting profiles developed from a common range-data using the various estimation processes. The range truth used here is generated using a known $\mathbf{m_x}$ and $\mathbf{\Lambda_x}$ with $\Pr(A) = 0.2$ and $\Delta R = 1000$m as in the previous cases. $\mathbf{m_x}$ is actually obtained from the mean values of $\hat{\mathbf{x}}_{PML}$ for $P = 64$ from the 500 trials of EM iteration used for evaluating the performance in Sec. 4.5. Fig. 6-11 plots $r_q = (\mathbf{H}_P \mathbf{m_x})_q$ vs. q. The covariance $\mathbf{\Lambda_x}$ used is a diagonal matrix with the first 32 components ranging from 1–6 and the last 32 components very small. Fig. 6-12 shows the mean and standard deviation $\sigma_{\mathbf{r}_q} = \sqrt{\text{diag}[\mathbf{H}_P \mathbf{\Lambda_x} \mathbf{H}_P^T]_q}$.

Both the piecewise-smoothed ML and the MAP estimated profiles are better ap-

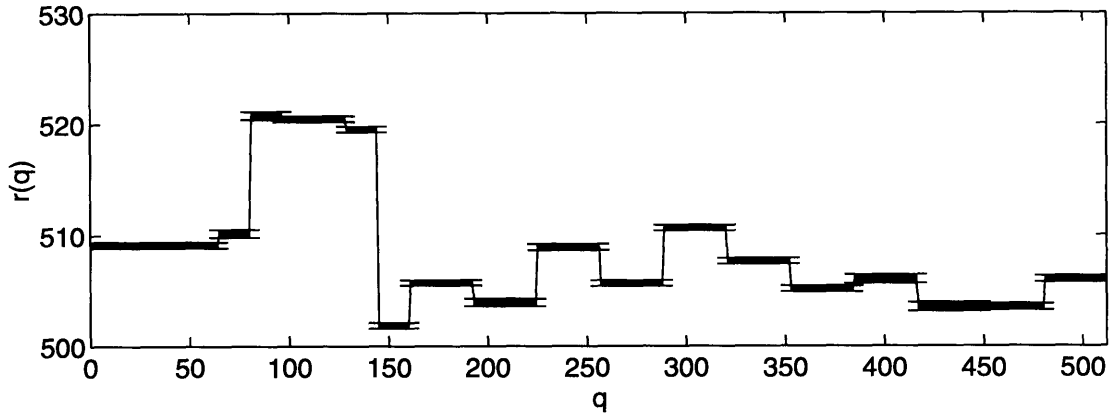

Figure 6-12: Mean range-truth $r_q = (\mathbf{H}_P \mathbf{m}_\mathbf{x})_q$ plotted with $\pm$ one standard deviation σ_{r_q} bounds.

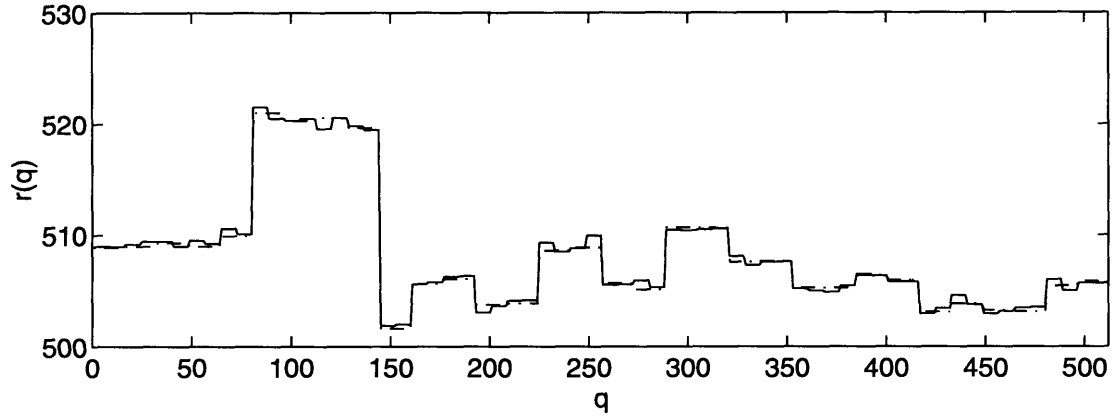

Figure 6-13: ML estimated profile at $P = 64$; dash-dotted line is the range-truth.

proximations of the original range profile than the ML estimate. However, the MAP estimate provides the closest approximation to the range-truth.

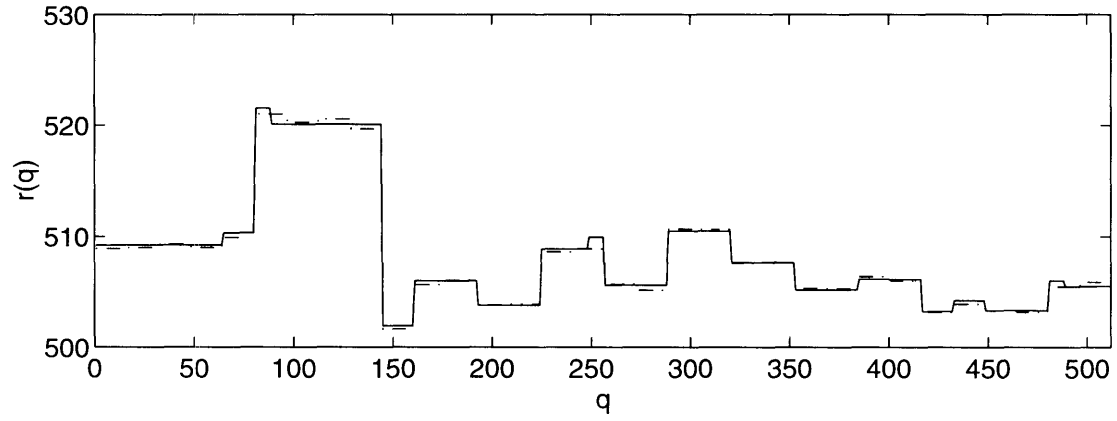

Figure 6-14: ML estimated profile after smoothing; dash-dotted line is the range-truth.

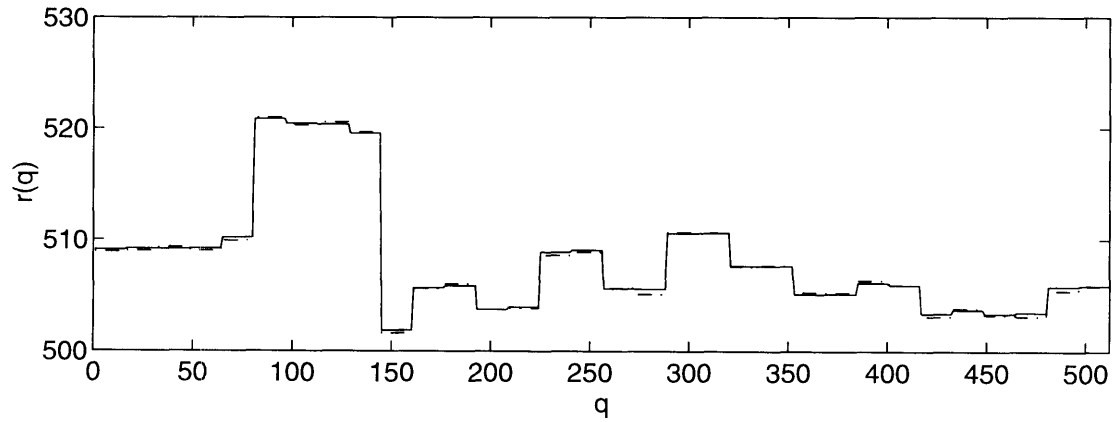

Figure 6-15: MAP estimated profile at $P = 64$; dash-dotted line is the range-truth.

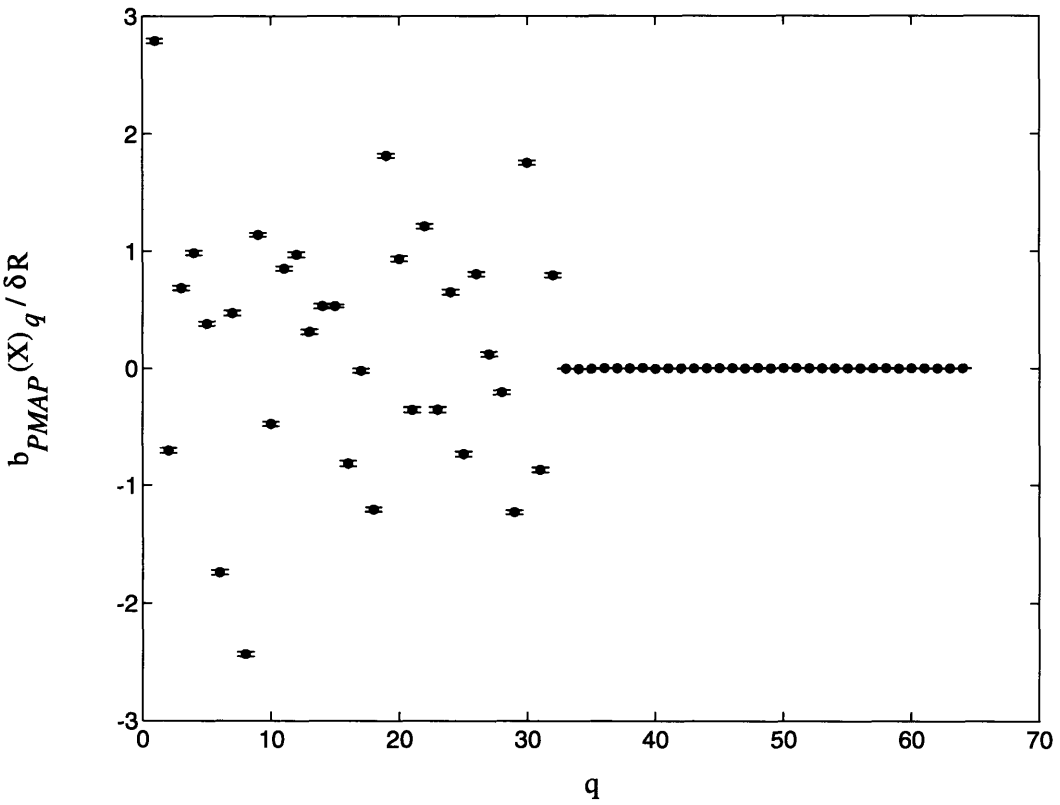

Figure 6-16: Normalized bias of the MAP estimate for $P = 64$.

The performance of the MAP estimate can be seen in Figs. 6-16 and 6-17. The bias of the coarse-scale parameters are larger than those for the ML and smoothed ML estimates. This is what we expected from the MAP estimation since the bias depends on the prior statistics of $\mathbf{X}_P$ and the actual value of the parameter itself. However, the MAP estimate has lower RMS errors than both the ML and the smoothed ML estimates due to the fact that we have additional information available from the prior statistics of $\mathbf{X}_P$. Based on this MAP approach, we can improve the performance of our system by directing the multiresolution range-profiler to target the specific kind of features we are interested in.

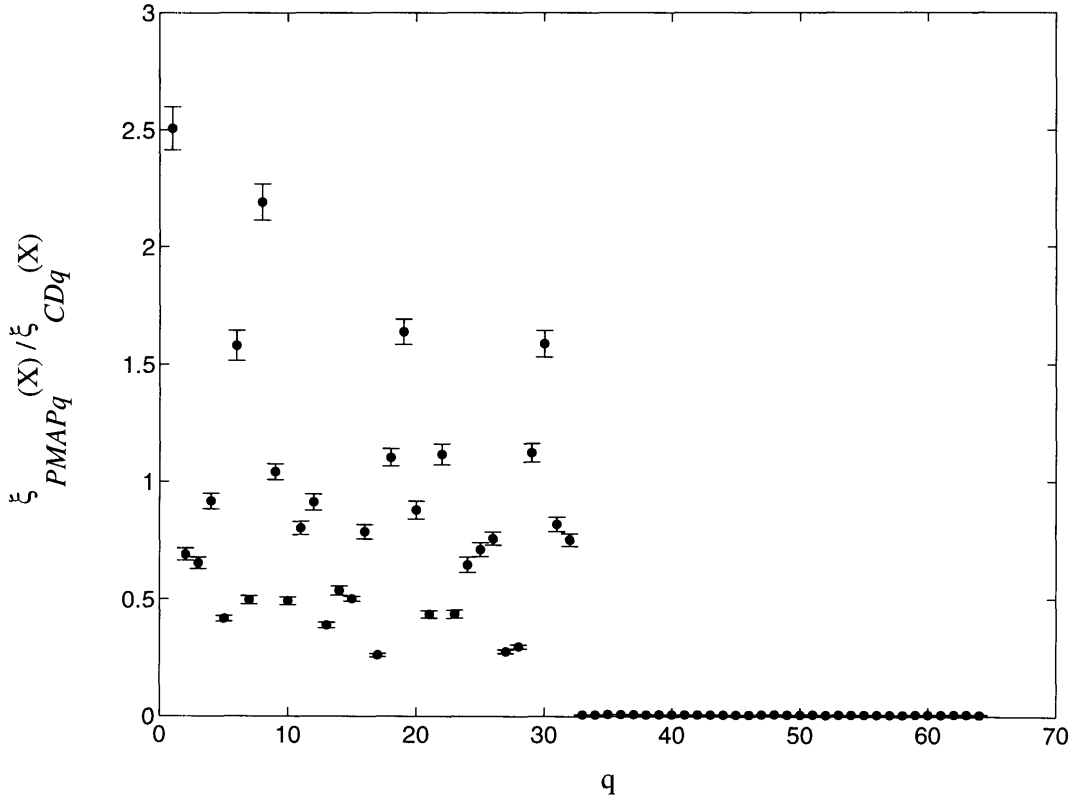

Figure 6-17: Normalized RMS error of the MAP estimate for $P = 64$.

Chapter 7

Conclusion

Coherent laser radar systems are capable of generating range images by raster scanning a field of view. These range images are subject to fluctuations arising from the combined effects of the laser speckle and local-oscillator shot noise. Because of this, a statistical approach to laser radar image processing seems most appropriate. Numerous research efforts have spun off ranging from the fundamentals of single-pixel statistics to 3-D object recognition and detection. The objective of this thesis is to continue this development by introducing the notion of multiresolution (wavelet) signal processing into the range profiling arena.

Building on previous work on planar-range profiling, we derived a more general theory of parametric maximum-likelihood range-profile estimation. Due to the presence of range anomalies, this estimation problem is inherently nonlinear in nature. We found that the expectation-maximization algorithm provided a computationally simple procedure for finding the ML estimate. The reliability of the ML estimate depended heavily on the quality of the initial estimate used to seed the EM algorithm. Satisfactory results were obtained using the wavelet-fitted range truth as the seed in the simulations. However, for practical realization of this estimation scheme, the recursive EM algorithm was an extremely effective approach despite its computional intensity.

Using the Haar-wavelet as an orthonormal basis, we developed a framework of multiresolution range profiling. Arranging the wavelets in some increasing fine-scale

behavior, we obtained a sequence of estimates with successively coarse-to-fine features extracted from the original range profile. The weights associated with the EM iterations are shown to provide a reliable indicator for terminating the coarse-to-fine estimation progression. Moreover, at the weight-determined stopping point, the ML estimate was approximately unbiased and had an error-covariance matrix approaching the complete-data bound.

For range profiles that comprise features at a variety of scales, proceeding to a higher resolution meant incorporating some extraneous fine-scale variations along the piecewise-constant areas. Extra parameters were used to represent these variations due to the fixed increments in resolution. Thus we needed a smoothing routine to suppress these variations in a spatially-varying manner in order to yield a minimal-dimensionality wavelet-fitted profile comprising the maximum resolution capability implied by the original image's single-pixel statistics. We found that an ad hoc process, which we called the piecewise smoothing method, was straightforward and extremely efficient in smoothing the estimated profile. However, this method is also shown to be biased towards a class of profiles with wide features and sharp discontinuities and against those with slowly-varying trends.

Another smoothing approach was to perform the ML estimation based only on the nonzero parameters implied by the wavelet-fitted range truth. The performance of this scheme was actually better than the ML estimate at the "optimal" resolution, since we were estimating the exact dimensions of the profile. We found that the availability of some prior knowledge of the parameter-vector can improve the smoothness of the estimated profile. This led to the development of the maximum a posteriori probability (MAP) estimation. Assuming the parameter-vector to be Gaussian, the MAP estimator was derived and implemented by slightly modifying the EM algorithm. Simulation results showed that the MAP indeed have better performance than the ML estimate if the prior statistics of the parameter-vector were relevant.

This thesis presented the basic framework of the multiresolution range-profiler for laser radars. There are certainly many directions in which future research work can pursue. The Haar-wavelet basis was chosen to provide an initial understanding of

the system. For more general and complicated scenarios, other wavelet-bases, both separable and non-separable, can be used. The Fourier series may have some interesting applications for certain models. In addition, the weight-determined termination criterion can be replaced with more sophisticated methods for specific applications. Real laser radar images can be used to test the estimation derived and implemented in this thesis. Finally, this method of multiresolution image processing can be applied to other sensors, such as peak-detecting Doppler imagers, FLIR models, and SAR and ISAR models.

Appendix A

Invertibility of $\mathbf{H}_P^T\mathbf{W}_P\mathbf{H}_P$

Here we will discuss the invertibility of the matrix $\mathbf{H}_P^T\mathbf{W}_P\mathbf{H}_P$ in Eq. 3.33 when $P < Q$. For $\Pr(A) < 1$, we have $w_q > 0$ for $1 \leq q \leq P$ and so the matrix $\mathbf{W}_P$ will be positive definite. Thus the required inverse exists if and only if $rank(\mathbf{H}_P) = P$. But this condition is guaranteed by the orthogonal matrix structure of $\mathbf{H}$.

In theory, the inverse of the matrix should always exist. In practice, computational underflow of $w_q(n)$ occurs when the range measurement is many δR away from the current estimated range value, i.e., the pixel appears to be anomalous based on the current estimated profile. From our termination criterion, we must have

$$\mathbf{W}_P(n) = \text{diag}[w_q(n)], \tag{A.1}$$

where

$$w_q(n) \simeq z_q = \begin{cases} 0 & \text{if pixel is presumed anomalous,} \\ 1 & \text{if pixel is presumed not anomalous.} \end{cases} \tag{A.2}$$

Since we know $\mathbf{H}_P$ has full rank by construction, we only need $\mathbf{Z}_P(n) = \text{diag}[z_q(n)]$ to span the range space of $\mathbf{H}_P^T$ for $\mathbf{H}_P^T\mathbf{W}_P\mathbf{H}_P$ to be numerically invertible. In essence, if $\tilde{R}_q \equiv R_q z_q$, we need $(\mathbf{H}_P^T\tilde{\mathbf{R}})_q$ to be nonzero for $1 \leq q \leq P$ to ensure numerical invertibility of $\mathbf{H}_P^T\mathbf{W}_P\mathbf{H}_P$. The REM algorithm is relavitively immune to such numerical difficulties because only a few pixels are treated as anomalous at each iteration.

Appendix B

Monotonicity of the EM Sequence

In the following we will prove an essential property of the EM algorithm: it produces a sequence of parameter-vector estimates, $\{\hat{\mathbf{x}}(n) : n = 0, 1, 2, \cdots\}$, whose associated likelihood sequence, $\{p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X}_P(n)) : n = 0, 1, 2, \cdots\}$, is monotonically increasing.

First, we define a companion function to $Q(\mathbf{X'}_P \mid \mathbf{X}_P)$ [16],

$$H(\mathbf{X'}_P \mid \mathbf{X}_P) \equiv \sum_{\mathbf{A}} \Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X}_P) \times \ln[\Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X'}_P)]. \tag{B.1}$$

It is then not difficult to verify that the incomplete data log-likelihood can be written as

$$L(\mathbf{X'}_P) \equiv \ln p_{\mathbf{r}|\mathbf{x}_P}(\mathbf{R} \mid \mathbf{X'}_P) = Q(\mathbf{X'} \mid \mathbf{X}_P) - H(\mathbf{X'}_P \mid \mathbf{X}_P). \tag{B.2}$$

Since a sequence of increasing likelihoods is the equivalent to a sequence of increasing log-likelihoods, we only need to show that

$$L[\hat{\mathbf{x}}_P(n+1)] - L[\hat{\mathbf{x}}_P(n)] > 0, \qquad \text{for} \quad n = 0, 1, 2, \cdots. \tag{B.3}$$

Following Eq. B.2, we have

$$L[\hat{\mathbf{x}}_P(n+1)] - L[\hat{\mathbf{x}}_P(n)] = \{Q[\hat{\mathbf{x}}_P(n+1) \mid \hat{\mathbf{x}}_P(n)] - Q[\hat{\mathbf{x}}_P(n) \mid \hat{\mathbf{x}}_P(n)]\} - \{H[\hat{\mathbf{x}}_P(n+1) \mid \hat{\mathbf{x}}_P(n)] - H[\hat{\mathbf{x}}_P(n) \mid \hat{\mathbf{x}}_P(n)]\}. \quad \text{(B.4)}$$

The maximization step in the EM update procedure makes $\{Q[\hat{\mathbf{x}}_P(n+1) \mid \hat{\mathbf{x}}_P(n) : n = 0, 1, 2, \cdots\}$ a non-decreasing sequence by construction, making the first term on the right of Eq. A.4 always positive. Thus our proof will be complete with the following simple lemma,

$$H(\mathbf{X'}_P \mid \mathbf{X}_P) \leq H(\mathbf{X}_P \mid \mathbf{X}_P). \quad \text{(B.5)}$$

Moreover, establishing the above equation is straightforward:

$$\begin{aligned} & H(\mathbf{X'}_P \mid \mathbf{X}_P) - H(\mathbf{X}_P \mid \mathbf{X}_P) = \\ & \quad \sum_{\mathbf{A}} \Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X}_P) \ln \left(\frac{\Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X'}_P)}{\Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X}_P)} \right) \\ & \leq \sum_{\mathbf{A}} \Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X}_P) \left(\frac{\Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X'}_P)}{\Pr(\mathbf{a} = \mathbf{A} \mid \mathbf{r} = \mathbf{R}, \mathbf{x}_P = \mathbf{X}_P)} - 1 \right) \\ & = 0, \end{aligned} \quad \text{(B.6)}$$

where the inequality follows from $\ln(z) \leq z - 1$ for $z > 0$.

Appendix C

Notes on Implementation

Below are some notes on the implementation of the multiresolution range-profiling simulation program.

C.1 Environments and Languages

The simulation code was implemented using C++, according to the AT&T 2.0 specification. It has been tested and built on both MS-DOS and UNIX environments. Microsoft Visual C++ was used to build the code on MS-DOS. Gnu C++ (g++) and AT&T C-fronts were used to build the code on UNIX. The actual simulations were carried out on DEC MIPS station and Cray Supercomputer.

C.2 Structures

There are two parts to the simulation program. The first part is a matrix manipulation package. The second part consists of the actual simulation routines. The Matrix classes have been heavily optimized to reduce the computational time. It is hoped that future projects could find some use in these optimized codes.

C.3 Optimizations

Many operations on Matrix have been heavily optimized. In particular, 'multiply', 'transpose' and 'solve' were made much faster by flattening out the abstract objects to direct pointers. Much fine-tuning was done to take advantage of the underlying architecture of both DEC MIPS and the Cray. Besides, the Matrix package was also designed with a constrained system in mind. In particular, it runs also on 16-bit MS-DOS environment. A DOS-extender (such as those by Phar Laps) is still needed for larger simulations though.

C.4 Performance

It takes 140 seconds to perform one trial of ML (REM) estimation with profile length 512 and resolution 32 on a DEC 5000/133. If the resolution is raised up to 64, the time requirement goes up to 400 seconds. The simulation runs around 25 times faster on a Cray X-MP EA/464.

C.5 Further optimizations

Any improvements in solving or multiplying matrices will lead to great gain in performance because these two operations were shown to be the bottlenecks of all computations. Their influence are even more drastic for larger matrices since these two operations take N-square time to perform.

References

[1] R.C. Harney and R.J. Hull, "Compact infrared radar technology," Proc. SPIE **227**, 162-170 (1980).

[2] A.B. Gschwendtner, R.C. Harney, and R.J. Hull, "Coherent IR radar technology," in D.K. Killinger and A. Mooradian, eds., *Optical and Laser Remote Sensing* (Springer-Verlag, Berlin, 1983).

[3] J.W. Goodman, "Statistical properties of laser speckle," in J.C. Dainty, ed., *Laser Speckle and Related Phenomena* (Springer-Verlag, Berlin, 1975).

[4] R.H. Kingston, *Detection of Optical and Infrared Radiation* (Springer-Verlag, Berlin, 1978), Chap. 3.

[5] J.H. Shapiro, R.W. Reinhold, and D. Park, "Performance analyses for peak-detecting laser radars," Proc. SPIE **663**, 38-56 (1986).

[6] J.H. Shapiro, "Target reflectivity theory for coherent laser radars," *Appl. Opt.* **21**, 3398-3407 (1982).

[7] M.B. Mark and J.H.Shapiro, "Multipixel, multidimensional laser radar system performance," Proc. SPIE **783**, 109-122 (1987).

[8] S.M. Hannon and J.H. Shapiro, "Laser radar target detection with a multipixel joint range-intensity processor," Proc. SPIE **999**, 162-175 (1988).

[9] S.M Hannon and J.H. Shapiro, "Active-passive detection of multipixel targets," Proc. SPIE **1222**, 2-23 (1990).

[10] T.J. Green, Jr. and J.H. Shapiro, and M.M. Menon, "Target detection performance using 3-D laser radar images," Proc. SPIE **1471**, 328-341 (1991).

[11] T.J. Green, Jr., "Three-Dimensional Object Recognition Using Laser Radar," Ph.D. thesis, Dept. of Elect. Eng. and Comput. Sci., MIT, 1992.

[12] T.J. Green, Jr., J.H. Shapiro, "Maximum-likelihood laser radar range profiling with the expectation-maximization algorithm," *Opt. Eng.* **31**, 2343-2354 (1992).

[13] S.G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Machine Intell. **PAMI-11**, 674-693 (1989).

[14] S.G. Mallat, "Multifrequency channel decompositions of images and wavelet models," IEEE Trans. Acoust,. Speech, Signal Processing **ASSP-37**, 2091-2110 (1989).

[15] O. Rioul and M. Vetterli, "Wavelets and signal processing," IEEE Signal Processing Magazine, 14-38 (Oct. 1991).

[16] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Ser.* **39**, 1-38 (1977).

[17] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley, New York, 1968.

[18] J.H. Shapiro and A.S. Willsky, *Stochastic Processes, Detection and Estimation Class Notes, MIT Subject 6.432*, 1991.

[19] A.W. Drake, *Fundamentals of Applied Probability Theory*, McGraw Hill, 1967.