THECRAMERRAOBOUNDFOR DISCRETE-TIME EDGE POSITION ESTIMATION

Alan Gatherer

Technical Report: CSL-TR-93-560

February 1993

This work was partially funded by the SRC Contract No. 9 l-MC-5 15 and the National Science Foundation, NSF Contact No. MIPS 8957058.

The Cramer Rao Bound for Discrete-time Edge Position Estimation

Alan Gatherer

Technical Report: CSL-TR-93-560

February, 1993

Computer Systems Laboratory
Departments of Electrical Engineering and Computer Science
Stanford University
Stanford, CA 943054055

Abstract

The problem of estimating the position of an edge from a series of samples often occurs in the fields of machine vision and signal processing. It is therefore of interest to assess the accuracy of any estimation algorithm. Previous work in this area has produced bounds for the continuous time estimator. In this paper we derive a closed form for the minimum variance bound (or Cramer Rao bound) for estimating the position of an arbitrarily shaped edge in white Gaussian noise for the discrete samples case. We quantify the effects of the sampling rate, the bandwidth of the edge, the shape of the edge and the size of the observation window on the variance of the estimator. We describe a maximum likelihood estimator and show that in practice this estimator requires fewer computations than standard correlation.

Key Words and Phrases: Edge detection, Cramer-Rao Bound, Pattern Recognition, Estimation Theory

---

This work partially funded by the SRC Contract No. 9 1 -MC-5 15 and the National Science Foundation, NSF Contract No. MIPS 8957058.

1.0 Introduction

Estimating the position of an edge from a finite number of samples of an image is a common signal processing operation in many fields including lithographic alignment in IC fabrication [1][6], pattern recognition and analysis [2], stereo vision [3], motion estimation [4] and aerial images [9]. It is important to know the relative accuracy of an edge-position estimation algorithm compared to what is achievable. If the variance of the estimation error for an edge-position estimation algorithm can be found, then it can be compared to the variance of other algorithms, but this does not tell the designer if another algorithm exists, possibly undiscovered, that performs better. What is needed is a lower bound on the estimation error variance of all possible estimation algorithms. In this paper we apply the Cramer Rao Bound (CRB) [5] to obtain a lower bound on the estimation error variance of *all* unbiased edge-position estimators, from a finite number of samples of an arbitrarily-shaped edge in additive white Gaussian noise. The derivation of the CRB uses a similar methodology as appeared in [6] where the CRB was derived for the case of pulses in noise. In a paper recently published in PAM1 [7], a closed form expression was presented for the CRB for edge position estimation of an edge passed through a continuous time Gaussian filter and then corrupted by white Gaussian noise. The results presented in this paper are more general compared to [7] in three ways. First, we use the more realistic discrete sampled data model, as all estimation algorithms are computed by computers or digital signal processors on digitized images. Another disadvantage of using the continuous time model is that it requires a measure of the power of the white Gaussian noise, which is difficult to obtain due to finite bandwidth constraints. We show that the CRB becomes independent of sampling position when the sampling rate is above the Nyquist rate of the edge. Second, our result holds for arbitrary edge shapes so that we can quantify the degree to which some edges can be more accurately estimated than others. Third, we quantify the effect of a finite observation window on the CRB and provide a simple rule of thumb for finding the minimum observation window size such that the effect of the finite window size on the CRB is kept within specified bounds.

The results contained in this paper apply to the estimation of the position of an edge in white Gaussian noise when the edge shape is known (though the amplitude may be unknown), the edge is known to occur within a certain observation window and no other features present within the window. The results we present are applicable in situations where the presence of a single edge has been detected and a high accuracy estimate of its position is required. This is a more constrained problem than that presented in [8] where an unknown number of edges are present. It has been stated [7] that the continuous time version of the Canny edge detector, when optimized for the Gaussian edge shape [8], does not asymptotically achieve

the CRB for the Gaussian edge shape. However, as the Canny edge detector is optimized to trade off estimation error and false edge detection, one would not expect it to minimize the estimation error for the situation of a exactly one edge in noise. The minimum mean square estimation error estimator for exactly one edge in noise is the maximum likelihood (ML) estimator. Kakarala and Hero [7] proposed an approximation to the ML estimator when the edge position was near the origin. We will show that the ML estimator can be implemented so that it has low complexity for edge estimation and in certain cases will be of lower complexity than standard correlation or filtering.

2.0 Problem Statement

The problem is to estimate the position of an edge whose shape is described by a function of unknown amplitude, corrupted by white Gaussian noise. We assume that the edge profile has been sampled with period T and that these samples are used to obtain an unbiased estimate of the edge location. The mathematical model is a continuous time signal

$$x(t) = as(t-\theta) + n(t), \tag{1}$$

where $s(.)$ is the edge shape which, without loss of generality, is assumed to be of unity height, $n(t)$ is a white Gaussian noise process, and a and θ are the unknown constants representing the amplitude and position of the edge respectively. Sampling over a window of length NT gives a vector $\boldsymbol{x}$ of N samples such that

$$\boldsymbol{x} = a\boldsymbol{s}_\theta + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{n}$ is a length-N vector of white Gaussian samples with variance σ_T^2, which will generally depend on the sample period T and the method of sampling. The vector $\boldsymbol{s}_\theta$ is a length-N vector whose i^{th} element iss $(iT-\theta)$.

3.0 The CRB for Unbiased Estimators of Edge Position

In this paper we shall consider the CRB for two important situations in edge detection. First we consider the case where the position and amplitude are both unknown constants with no associated probability distribution. Second we examine the case where only the position is unknown. We will highlight the similarities and differences in these two situations.

3.1 The CRB for Unknown Nonrandom Amplitude and Position

If there are no known probability density functions for the amplitude and position then they must be treated as unknown constants. We shall now derive the Cramer Rao Bound (CRB) for this case. The general definition of the CRB is stated in [5]. We give it here to show the notation used in this paper.

Theorem *3.1. Consider a system whose output* $\boldsymbol{x}$ *has statistics defined by an unknown, nonrandom, length-M parameter vector* $\boldsymbol{q}$. *If for a given instance of* $\boldsymbol{x}$ *an unbiased estimate* $\hat{\boldsymbol{q}}$ *of* $\boldsymbol{q}$ *is obtained, then the variance of the estimation error* $(\hat{q}_i - q_i)$ *of the* i^{th} *element of* q *is lower bounded by the* ii^{th} *element of the* $M \times M$ *matrix* F^{-1}, *whose inverse is given by*

$$F = E\left[\left(\frac{\partial \ln (p(\boldsymbol{x}|\boldsymbol{q}))}{\partial \boldsymbol{q}}\right)\left(\frac{\partial \ln (p(\boldsymbol{x}|\boldsymbol{q}))}{\partial \boldsymbol{q}}\right)^{*}\right]. \tag{3}$$

The partial derivative in (3) is the column vector of the derivative of $p(\boldsymbol{x}|\boldsymbol{q})$ *with respect to the elements of* q. *The function* $p(\boldsymbol{x}|q)$ *is the probability of* $\boldsymbol{x}$ *given* q. $\boldsymbol{F}$ *is called the Fisher information matrix.*

By nonrandom parameters, we mean that the parameters are unknown constants and have no associated probability density.

In our case the unknown parameter vector is $\boldsymbol{q} = [\theta, a]$ and

$$p(\boldsymbol{x}|\boldsymbol{q}) = \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left(-\frac{(\boldsymbol{x} - a\boldsymbol{s}_\theta)^{*}(\boldsymbol{x} - a\boldsymbol{s},)}{2\sigma_T^2}\right). \tag{4}$$

So the log likelihood is

$$\ln (p(\boldsymbol{x}|q)) = -\frac{1}{2}\ln (2\pi\sigma_T^2) - \frac{(\boldsymbol{x} - a\boldsymbol{s},)^{*}(\boldsymbol{x} - a\boldsymbol{s}_\theta)}{2\sigma_T^2}. \tag{5}$$

Define the vector $d_\theta = \frac{\partial}{\partial\theta}s_\theta$, which is the sample vector of $-s'(t-\theta)$, then

$$\frac{\partial}{\partial\theta}\ln (P(x|\boldsymbol{q})) = \frac{(\boldsymbol{x} - a\boldsymbol{s},)^{*}\boldsymbol{d}_\theta a}{\sigma_T^2} \tag{6}$$

and

$$\frac{\partial}{\partial a}\ln\left(p\left(\boldsymbol{x}|\,\boldsymbol{q}\right)\right) = \frac{\left(\boldsymbol{x}-a\boldsymbol{s}_\theta\right)^{*}\boldsymbol{s}_\theta}{\sigma_T^2}\ . \tag{7}$$

Defining the matrix $P = [\,\boldsymbol{d}_\theta a, \boldsymbol{s}_\theta]$ and noting that $(\mathrm{x} - \boldsymbol{as},) = \boldsymbol{n},\ \boldsymbol{we}$ get

$$\frac{\partial \ln\left(p\left(\boldsymbol{x}|\,\boldsymbol{q}\right)\right)}{\partial \boldsymbol{q}} = \sigma_T^{-2}P^{*}\boldsymbol{n}\,. \tag{8}$$

We can now state the following theorem:

Theorem *3.2. For an edge estimation problem as described in (2) with parameter vector* $4 = [\theta, \boldsymbol{a}]$ *, the Fisher information matrix is*

$$F = \sigma_T^{-2}\begin{bmatrix} a^2\|\boldsymbol{d}_\theta\|^2 & a\boldsymbol{d}_\theta^{*}\boldsymbol{s}_\theta \\ a\boldsymbol{d}_\theta^{*}\boldsymbol{s}_\theta & \|\boldsymbol{s}_\theta\|^2 \end{bmatrix}. \tag{9}$$

Proof: Using equations (3) and (8),

$$F = \sigma_T^{-4}P^{*}E\left[\boldsymbol{n}\boldsymbol{n}^{*}\right]P\ , \tag{10}$$

but as the noise is white, the covariance of the noise is $I\sigma_T^2$, so that

$$F = \sigma_T^{-2}P^{*}P. \tag{11}$$

From the definition of *P,* (9) follows.

To find the CRB we take the inverse of (9) to obtain

$$E\left[\left(\theta-\hat{\theta}\right)^2\right] \geq \frac{\sigma^2{}_T\|\boldsymbol{s}_\theta\|^2}{a^2\|\boldsymbol{d}_\theta\|^2\|\boldsymbol{s}_\theta\|^2 - a^2\left(\boldsymbol{d}_\theta^{*}\boldsymbol{s}_\theta\right)^2}. \tag{12}$$

The inner product $\boldsymbol{d}_\theta^{*}\boldsymbol{s}_\theta$ can be described as $\boldsymbol{d}_\theta^{*}\boldsymbol{s}_\theta = \|\boldsymbol{d}_\theta\|\|\boldsymbol{s}_\theta\|\cos\alpha$, where α is the angle between the N dimensional vectors $\boldsymbol{d}_\theta$ and $\boldsymbol{s}_\theta$, giving

$$E\left[\left(\theta-\hat{\theta}\right)^2\right] \geq \frac{\left(\sigma_T/a\right)^2}{\left|\boldsymbol{d}_\theta\right|^2\sin^2\alpha}\ . \tag{13}$$

The estimation error is heavily dependent on the angle between the edge shape and its derivative. The angle between these two vectors can be thought of as defining the degree to which the amplitude being unknown effects the position estimate.

3.2 The CRB for Known Amplitude and Unknown Position

Suppose the edge amplitude, ***a,*** is known. Then the CRB reduces to a single parameter problem:

$$E\left[(\theta - \hat{\theta})^2 \right] \geq \left(E\left[\left(\frac{\partial}{\partial\theta} \ln \left(p\left(\boldsymbol{x}|\,\theta\right)\right)\right)^2 \right]\right)^{-1}$$

$$= \left(E\left[\left(\frac{(\boldsymbol{x} - a\boldsymbol{s}_\theta)^* \boldsymbol{d}_\theta a}{\sigma_T^2} \right)^2 \right] \right)^{-1}$$

$$= \frac{(\sigma_T / a)^2}{\|\boldsymbol{d}_\theta\|^2} \tag{14}$$

Comparing this with equation (13) we see that the CRB of the position estimate for an edge of unknown amplitude is equal to the CRB of the estimate of a known amplitude when the angle CC is $\pi/2$. The angle α defines the effect of the unknown amplitude on the position estimation. When α is $\pi/2$, knowledge of the amplitude has no effect on the position estimation. As α decreases, the position estimation error increases.

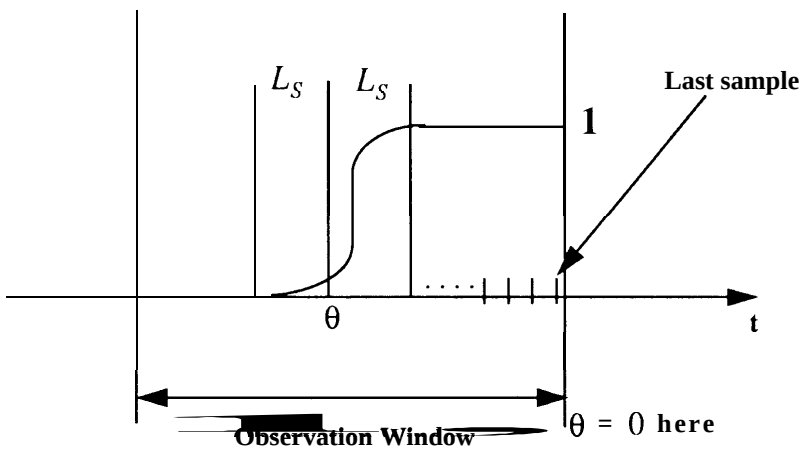
4.0 The CRB When Sampling Above the Edge Nyquist Rate

In general the CRB will be dependent on the sampling positions relative to θ because at low sampling rates the values of $\|\boldsymbol{d}_\theta\|^2, \|\boldsymbol{s}_\theta\|^2$ and $\boldsymbol{d}_\theta^* \boldsymbol{s}_\theta$ are dependent on the sampling positions relative to the edge position. This will cause the CRB to fluctuate with a periodicity equal to the sampling period. However, we will show that the CRB can be approximated by a sampling point independent expression under the conditions that the sampling rate is sufficiently high and the edge is contained within the observation window.

4.1 Assumptions

If an edge is contained within the window of observation, it must have finite duration and therefore it must have infinite bandwidth. Although it is possible to define an edge so that it has infinite duration (for example the Gaussian edge used in [7]), edges observed in real applications exhibit negligible change outside of a finite region. Physical constraints will also keep most of the energy of the edge within a finite frequency region. Therefore for sufficiently high sampling rate the aliasing effect that occurs will be small. We will assume that the energy of the aliasing spectrum is less than ε times the energy of the edge. The function s (t) only changes over a finite region. Therefore we will define the start of the edge as the furthest right point that has all values to the left of it exactly equal to zero. The end of the edge is similarly defined to be the left most point that has all values to the right of it exactly equal to unity. Without loss of generality, we will define s (t) such that the center of the edge is at the origin, the start of the edge occurs at a distance of L_S to the left of center and the end of the edge occurs at a distance of L_S to the right of center. The position of the edge θ will be measured from half a sample past the last sample on the right of the observation window (see Figure 1). As the edge is contained within the observation window, this will make the position negative. The function that equals s (t) in the region $t \in [-L_S, L_S]$ and is zero otherwise will be defined as $\tilde{s}$ (t) .

Figure 1. Conventions for edge position and size.



4.2 The Equivalence of Sums and Integrals

The power theorem [10] is a generalization of Parseval's theorem and relates inner products in time to those in frequency. Its continuous time form is

$$\int_{-\infty}^{\infty} h(t)\, g(t)^{*} dt = \int_{-\infty}^{\infty} H(f)\, G^{*}(f)\, df, \tag{15}$$

where $H(f)$ and $G(f)$ are the Fourier transforms of two real functions $h(t)$ and $g(t)$ respectively. For brevity we will adopt the bracket notation for integrals of products of functions, so that

$$\langle h, g \rangle = \int_{-\infty}^{\infty} h(t)\, g(t)^{*} dt \tag{16}$$

and equation (15) is equivalent to

$$\langle h, g \rangle = \langle H, G \rangle \tag{17}$$

The discrete-time power theorem is

$$\sum_{i=-\infty}^{\infty} h_i g_i^{*} = \int_{-1/2}^{1/2} \tilde{H}(f)\, \tilde{G}^{*}(f)\, df, \tag{18}$$

where $\tilde{H}(f)$ and $\tilde{G}(f)$ are the discrete-time Fourier transforms of the sample sequences of $h(t)$ and $g(t)$, denoted by h_i and g_i respectively. For sampling period T, we define there to be "&-small aliasing effect" on $H(f)$ when

$$\int_{1/2T}^{\infty} \| H(f) \|\, df < \varepsilon \sqrt{\int_{-1/2T}^{1/2T} \| H(f) \|^{2} df}\,. \tag{19}$$

This definition is motivated by the fact that aliasing is thc sum of multiple overlapping spectrum samples. It is therefore the area under the magnitude response above $f = 1/2T$ that determines the aliasing rather than the power above $\mathbf{f} = 1/2T$. With sampling period T,

$$\tilde{H}(f) = \frac{1}{T} \left(H\left(\frac{f}{T}\right) + \alpha_{H}\left(\frac{f}{T}\right) \right) \tag{20}$$

where

$$\alpha_{H}(f) = \sum_{i=-\infty,\, i \neq 0}^{\infty} H\left(f - \frac{i}{T}\right) \tag{21}$$

and there are identical definitions for $\boldsymbol{G}(f)$. The spectrums $\alpha_H(f)$ and $\alpha_G(f)$ are the aliasing spectrum. Assuming (19) holds then

$$\left(\int_{-1/2T}^{1/2T} \| \alpha_H(f) \|^2 df\right) \leq \left(\int_{-1/2T}^{1/2T} \sum_{i=-\infty, i\neq 0}^{\infty} \left\| H(f-\frac{i}{T}) \right\| df\right)^2$$

$$= \left(2 \int_{1/2T}^{\infty} \| H(f) \| df\right)^2$$

$$< 4\varepsilon^2 \int_{-1/2T}^{1/2T} \| H(f) \|^2 df. \tag{22}$$

In the same manner from (19) we obtain

$$\left(\int_{1/2T}^{\infty} \| H(f) \|^2 df\right) < \varepsilon^2 \int_{-1/2T}^{1/2T} \| H(f) \|^2 df. \tag{23}$$

Similarly, with E-small aliasing effect on $\boldsymbol{G}(f)$

$$\int_{-1/2T}^{1/2T} \| \alpha_G(f) \|^2 df < 4\varepsilon^2 \int_{-1/2T}^{1/2T} \| G(f) \|^2 df \tag{24}$$

and

$$\left(\int_{1/2T}^{\infty} \| G(f) \|^2 df\right) < \varepsilon^2 \int_{-1/2T}^{1/2T} \| G(f) \|^2 df \tag{25}$$

From (1 S), (20) and (21) and a change of variables WC obtain

$$\sum_{i=-\infty}^{\infty} h_i g_i = \frac{1}{T} \int_{-1/2T}^{1/2T} (H(f) G^*(f) + H(f) \alpha_G(f) + G^*(f) \alpha_H(f) + \alpha_H(f) \alpha_G(f)) df, \tag{26}$$

and from (22), (24), (26) and the Cauchy-Schwartz inequality we obtain the inequality

$$\left| \sum_{i=-\infty}^{\infty} h_i g_i - \frac{1}{T} \int_{-1/2T}^{1/2T} H(f) G^*(f) df \right| < 4(\varepsilon + \varepsilon^2) \sqrt{\int_{-1/2T}^{1/2T} \| G(f) \|^2 df \int_{-1/2T}^{1/2T} \| H(f) \|^2 df}. \tag{27}$$

Using (23), (25) and the Cauchy-Schwartz inequality we obtain

$$\left| \frac{1}{T} \int_{-1/2T}^{1/2T} H(f)\, G^*(f)\, df - \frac{1}{T} \int_{-1/2T}^{1/2T} H(f)\, G^*(f)\, df \right| < 2\varepsilon^2 \sqrt{\int_{-1/2T}^{1/2T} \| G(f) \|^2 df \int_{-1/2T}^{1/2T} \| H(f) \|^2 df}. \quad (28)$$

Finally, combining (27) and (28) we obtain

$$\left| \sum_{i=-\infty}^{\infty} h_i g_i - \frac{1}{T} \int_{-\infty}^{\infty} H(f)\, G^*(f)\, df \right| < (4\varepsilon + 6\varepsilon^2) \sqrt{\int_{-1/2T}^{1/2T} \| G(f) \|^2 df \int_{-1/2T}^{1/2T} \| H(f) \|^2 df} \quad . \quad (29)$$

Informally, (29) implies that for sampling rate high enough so that there is &-small aliasing on G (f) and $H(f)$ then

$$\sum_{i=-\infty}^{\infty} h_i g_i = \frac{1}{T} \int_{-\infty}^{\infty} H(f)\, G^*(f)\, df + O(\varepsilon). \quad (30)$$

The edge is of finite duration, therefore we can assume that it is completely contained within the observation window. If there is &-small aliasing of the derivative then, adopting the bracket notation as in (17), we obtain

$$\| \boldsymbol{d}_\theta \|^2 = \sum_{i=-\infty}^{\infty} s'(iT-\theta)^2 = \frac{1}{T} \langle s', s' \rangle (1 + O(\varepsilon)). \quad (31)$$

If there is also &-small aliasing of the derivative then

$$\boldsymbol{d}_\theta^* \boldsymbol{s}_\theta = \sum_{i=-\infty}^{\infty} s(iT-\theta)\, (-s'(iT-\theta)) = -\frac{1}{T} (\langle s', \tilde{s} \rangle + O(\varepsilon) \sqrt{\langle s, s \rangle \langle s', s' \rangle}). \quad (32)$$

Applying (30) to the calculation of $\| \boldsymbol{s}_\theta \|^2$ is not as straightforward as for $\| \boldsymbol{d}_\theta \|^2$ because the non-zero portion of the waveform is not contained within the window but instead is unity out to infinity on the upside of the edge. However, because of the way the origin was defined in Section 4.1, reflection of the window about the origin produces a pulse that has the same Nyquist rate as the edge and whose samples are those of the edge repeated twice. Therefore the sum of the pulse samples squared is equal to $2\| \boldsymbol{s}_\theta \|^2$. The integral of the left half of the pulse squared can be divided into two parts. The first is the integral over the edge region. This integral is the power in $\tilde{s}$ (t) . The second is the integral over the region of unit amplitude which has length $|\theta| - L_S$. Hence

$$\|s_\theta\|^2 = \frac{1}{T}\left(\langle \tilde{s}, \tilde{s}\rangle - L_S + |\theta|\right)\left(1 + O(\varepsilon)\right). \tag{33}$$

Combining (31), (32) and (33) we obtain

$$\|d_\theta\|^2\|s_\theta\|^2 - \left(d_\theta^* s_\theta\right)^2 = \frac{1}{T}\left(\langle s', s'\rangle\left(\langle \tilde{s}, \tilde{s}\rangle - L_S + |\theta|\right) - \langle s', \tilde{s}\rangle^2\right)\left(1 + 0(E)\right) \tag{34}$$

From equations (12), (33) and (34), for &-small aliasing of the edge and its derivative and with the edge contained within the observation window the CRB for unknown amplitude can be written as

$$\text{CRB} = \frac{T\sigma_T^2/a^2}{\langle s', s'\rangle\left(1 - \dfrac{\langle \tilde{s}, s'\rangle}{\langle s', s'\rangle\left(\langle \tilde{s}, \tilde{s}\rangle - L_S + |\theta|\right)}\right)}\left(1 + O(\varepsilon)\right), \tag{35}$$

The angle α is defined by

$$\cos^2\alpha = \frac{\langle \tilde{s}, s'\rangle}{\langle s', s'\rangle\left(\langle \tilde{s}, \tilde{s}\rangle - L_S + |\theta|\right)} \tag{36}$$

which defines α only in terms of the edge shape and θ. Similarly from (14) and (31) the CRB for known amplitude can be written as

$$\text{CRB} = \frac{T\sigma_T^2/a^2}{\langle s', s'\rangle}\left(1 + O(\varepsilon)\right). \tag{37}$$

In what follows we shall omit the $(1 + 0(E))$ for brevity and will say the CRB is approximately equal when it is equal $0(E)$.

4.3 The CRB for an Edge in Isolation

If we assume that the edge is isolated near the center of a large observation window then $|\theta| \gg L_S$. We shall also assume that the observation window is large enough so that $|\theta| \gg \langle \tilde{s}, \tilde{s}\rangle$ (Note that it is possible for $\langle \tilde{s}, \tilde{s}\rangle \gg L_S$ if there is significant overshoot). Therefore it follows that

$$\langle \tilde{s}, \tilde{s}\rangle - L_S + |\theta| \approx |\theta|. \tag{38}$$

Equation reduces (35) to

$$\text{CRB} \approx \frac{T\sigma_T^2/a^2}{\left(\langle s', s'\rangle - |\theta|^{-1}\langle \tilde{s}, s'\rangle^2\right)}. \tag{39}$$

Using the Cauchy-Schwartz inequality, (37), and $|\theta| \gg \langle \tilde{s}, \tilde{s} \rangle$, we obtain

$$|\theta|^{-1} \langle s', \tilde{s} \rangle^2 \leq |\theta|^{-1} \langle s', s' \rangle \langle \tilde{s}, \tilde{s} \rangle \ll \langle s', s' \rangle \tag{40}$$

Hence

$$\text{CRB} \approx \frac{T\sigma_T^2/a^2}{\langle s', s' \rangle}. \tag{41}$$

and the CRB for unknown amplitude becomes approximately equal to (37), the CRB for known amplitude. The angle α in this case is guaranteed to be close to $\pi/2$ because the uncertainty about the amplitude of the edge does not effect the position estimate, as the long tail of constant amplitude allows the amplitude to be estimated accurately within the data window.

4.4 The CRB for an Edge Not in Isolation

For the inequality (41) to be tight, we require that the edge is in isolation so that $|\theta| \gg L_S$ and $|\theta| \gg \langle \tilde{s}, \tilde{s} \rangle$. The "much greater than" symbol is not a practical guide to how big the observation window should be. If the edge whose position is to be estimated is close to other features then the size of the observation window will be limited and therefore the edge may not be in isolation. To minimize computation time, the window should be made as small as possible without effecting the accuracy of the result. Therefore in this section we examine the effect of finite window size on the CRB. We shall write (35) as follows

$$\text{CRB} \approx \frac{T\sigma_T^2/a^2}{\langle s', s') \ (1 - (\cos^2\tilde{\alpha})\ \gamma_\theta)}, \tag{42}$$

where

$$\cos^2\tilde{\alpha} = \frac{\langle s', \tilde{s} \rangle}{\langle s', s' \rangle \langle \tilde{s}, \tilde{s} \rangle} \tag{43}$$

and therefore $\tilde{\alpha}$ is the angle between $\tilde{s}$ (t) and s' (t) , and is only dependent on the edge shape. The term

$$\gamma_\theta = \frac{\langle \tilde{s}, \tilde{s} \rangle}{\langle \tilde{s}, \tilde{s} \rangle - L_S + |\theta|} \tag{44}$$

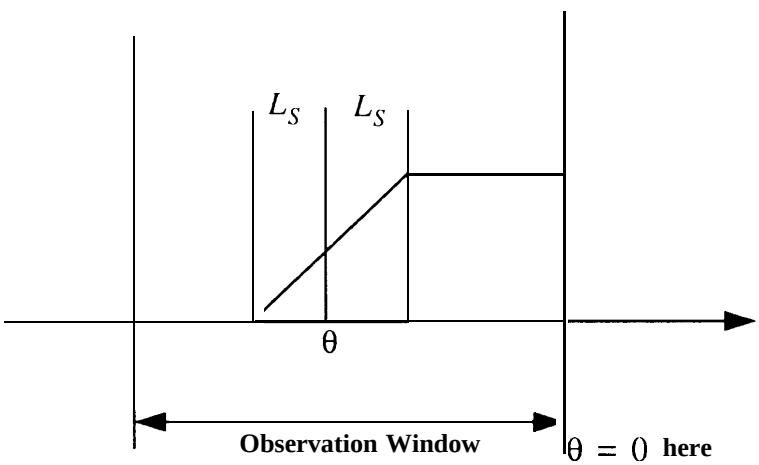
is an edge power dependent term that is monotonically decreasing in $|\theta|$ and defines the region over which the angle $\tilde{\alpha}$ has any effect on the CRB. Both γ_θ and $\tilde{\alpha}$ terms are edge-shape dependent but a simple upper bound on the CRB can be derived by noting that $\cos^2\tilde{\alpha} \le 1$. Therefore comparing (41) and (42), we get

$$\text{CRB}_{\text{isolation}} \le \text{CRB} \le \text{CRB}_{\text{isolation}} \left(\frac{\langle \tilde{s}, \tilde{s} \rangle + |\theta| - L_S}{|\theta| - L_S} \right), \tag{45}$$

where $\text{CRB}_{\text{isolation}}$ is the CRB with an infinitely large window. To guarantee that the fractional increase in the CRB due to finite window size is less than $1/p$, the window must extend beyond the edge by p times the energy of the edge (this can be shown by referring to the definitions of L_S and $|\theta|$ in Figure 1).

The above result is a simple, general formula for calculating an upper bound on the CRB irrespective of the edge shape. For a given edge shape a more accurate bound can be derived. As an example, we calculated the terms in (42) for the ramp edge shown in Figure 2. Note that there is slope discontinuity at the ends of the ramp edge. However, the ramp shape can be altered by an arbitrarily small amount within an arbitrary small region about the ramp edge to obtain a continuous derivative while all quantities of interest change by an arbitrarily small fraction. For the ramp edge we obtain $\cos\tilde{\alpha} = 3/4$ and $\gamma_\theta = 2L_S/(3|\theta| - L_S)$. This implies that for the ramp edge, to guarantee that the fractional increase in the CRB due to finite window size is less than $1/p$, the window must extend beyond the edge by $(p - 1/3)/4$ times the width of the edge. So for a less than 2% increase in the CRB due to finite window size, the window must extend past the edge by at least 12.42 times the length of the edge.

Figure 2. A ramp edge.



5.0 Maximum Likelihood Edge Estimation

The additive noise is assumed to be white Gaussian and therefore the Maximum Likelihood (ML) estimate is also the minimum mean-squared estimate [11]. For the purposes of verifying the CRB results we can use ML as a tool since ML estimation comes very close to the CRB for high SNR. We shall briefly describe the ML estimation method for the problem at hand.

5.1 Maximum Likelihood for Unknown Amplitude

The minimum mean-squared error estimates are the values $\hat{\theta}$ and $\hat{a}$ that minimize the function

$$f(\theta, a) = \| x - a s_\theta \|^2 . \tag{46}$$

Because this function is quadratic in a, there is a closed form solution for $\hat{a}$ given $\hat{\theta}$:

$$\hat{a} = \frac{s_\theta^* x}{\| s_\theta \|^2}, \tag{47}$$

Substituting (47) back into (46) gives

$$f(\theta) = \left\| \left(-\frac{s_\theta s_\theta^*}{\| s_\theta \|^2} \right) x \right\|^2 = x^* \left(I - \frac{s_\theta s_\theta^*}{\| s_\theta \|^2} \right) x, \tag{48}$$

where I is the identity matrix. The ML estimate $\hat{\theta}$ is therefore the minimum of a nonlinear, scalar function. Minimization can be performed using Newton's method [12]. As x is independent of θ, minimizing $f(\theta)$ is equivalent to maximizing the function

$$\tilde{f}(\theta) = \left(\frac{s_\theta^* x}{\| s_\theta \|} \right)^2 \tag{49}$$

which can be done to the nearest sample by standard correlation.

We now present a comparison of the ML mean-squared error versus the CRB for edge position estimation. Estimation of the ML mean-squared error is obtained by averaging the estimation error of 9000 simulation runs of a ML estimator. The ramp edge shown in Figure 2 was simulated with sampling period T equal to 1, the half edge width L_S equal to 10, the amplitude a equal to 1 and the noise power σ_T^2 equal to 0.0001. An example of this edge is shown in Figure 3. The small aliasing condition of Section 4.2 holds

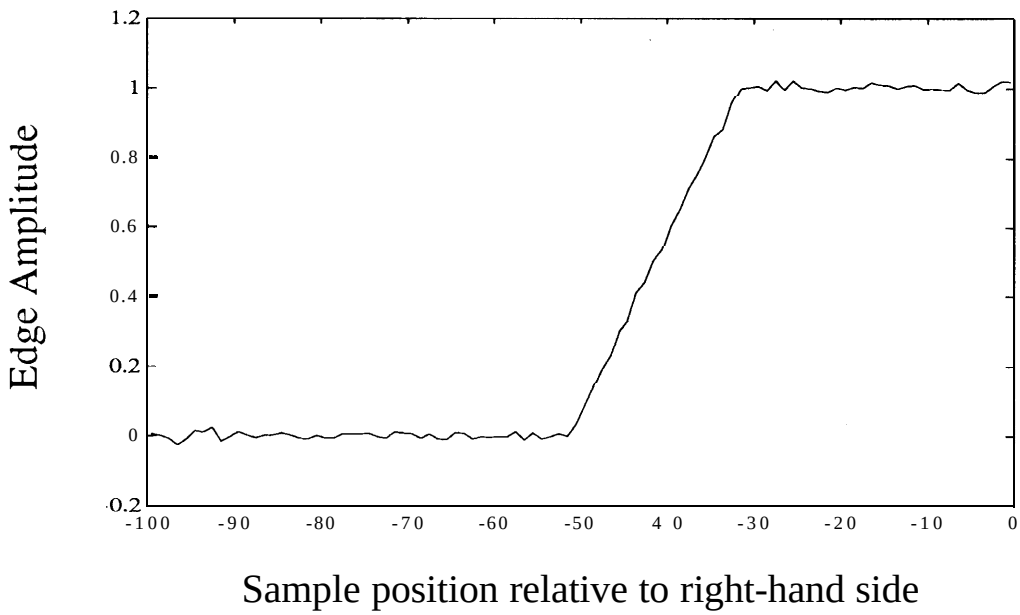
in this case with the power in the aliasing being less than -50dB compared to the signal power in the sampling bandwidth. The CRB for the ramp edge in Figure 2 can be shown to be

$$\text{CRB}_{\text{ramp}} = \frac{T\sigma_T^2/a^2}{\frac{1}{2L_S}\left(1 - \frac{3L_S}{2\,(3|\theta| - L_S)}\right)}. \tag{50}$$

For the ramp edge in Figure 3 the CRB becomes

$$\text{CRB}_{\text{ramp}} = \frac{2 \times 10^{-3}}{1 - \frac{15}{3|\theta| - 10}}. \tag{51}$$

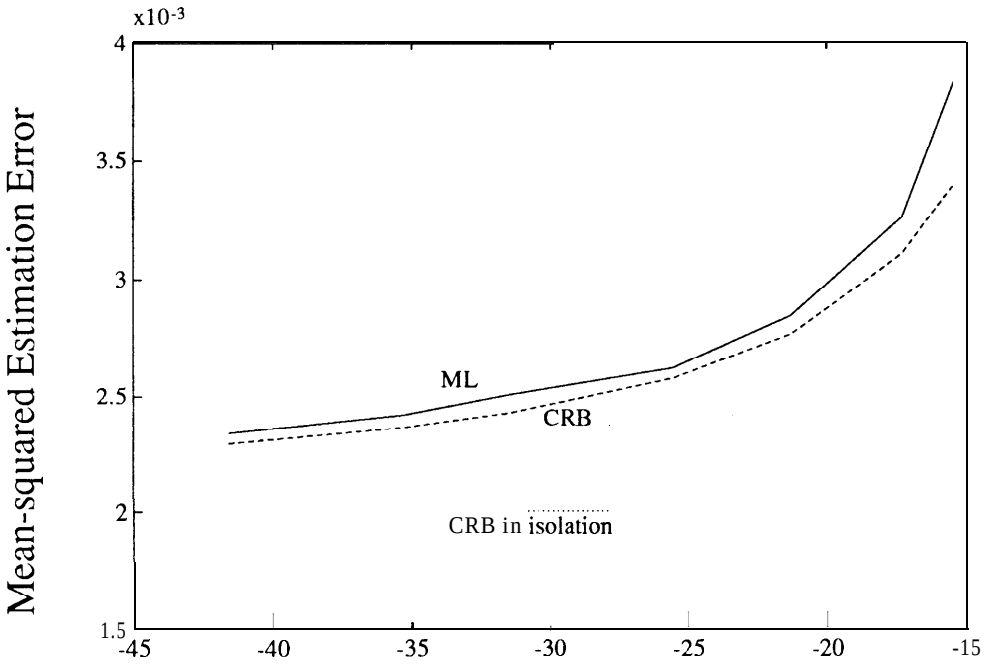
Figure 3. Ramp Edge with noise used in the numerical example.



To perform the ML estimation, it was first assumed that the cost function (48) was minimized using a modified Newton's method similar to that described in [13]. The details of this algorithm are given in Appendix A. The resulting ML mean-squared error is plotted for varying edge position θ in Figure 4. Notice that the ML estimate comes close to achieving the CRB so that the CRB is a tight bound on the minimum estimation error variance. The CRB decreases as the edge position moves away from the edge of the observation window. The CRB for this edge in isolation is also shown in Figure 4. The CRB can be used to make an

informed trade off between complexity and performance of the ML estimate. If a rough estimate of the position is obtained by some low complexity method then, using the CRB, a second window can be chosen around this estimate that is just large enough to obtain the ML estimate at the required accuracy therefore minimizing the computational complexity of the ML estimate.

Figure 4. ML estimation, unknown amplitude compared to the CRB versus edge position.



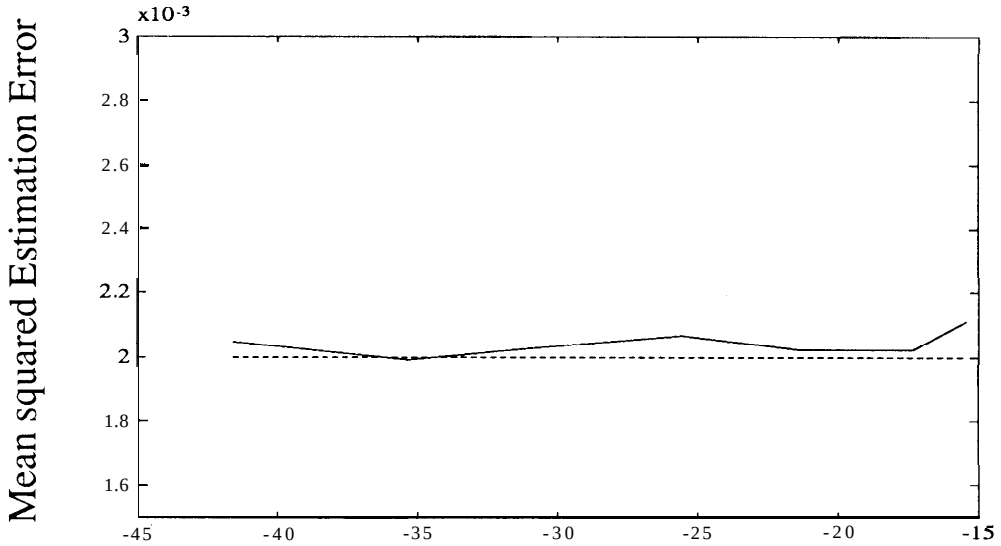
5.2 Maximum Likelihood for Known Amplitude

If the amplitude is known then the cost function changes accordingly. From equation (46) the function to be maximized is now

$$f(\theta) = x^* s_\theta - \frac{a}{2}\| s_\theta \|^2, \tag{52}$$

which can be maximized using Newton's method (details given in Appendix B). We repeat the experiment from Section 5.1 for the known amplitude case with the results shown in Figure 5. The performance now almost achieves the CRB in isolation independently of the position of the edge, as was expected. Though the algorithm's performance dips below the CRB at one point this well within the limits of experimental error.

Figure 5. ML estimation, known amplitude compared to the CRB versus edge position.



5.3 Computational Complexity of ML

In the examples presented in Sections 5.1 and 5.2 the Newton's algorithm was iterated 10 times. In our simulations the quadratic nature of the convergence of Newtons algorithm generally produces convergence in less than 10 iterations from starting states separated from the actual position by less than 200 samples. Therefore, for our example we assume that the ML converges in 10 steps when the edge is contained within a window of 256 samples. Appendix A and Appendix B discuss the computational complexity of these algorithms. For the known amplitude case, each iteration requires less than 22 multiply-accumulates (see Appendix B) so the ML estimation completes in 220 multiply-accumulates. Standard, nearest sample correlation requires two FFTs, one vector multiply, and a peak search. Even not counting the search this requires 4352 multiply-accumulates for a vector of length 256 samples. Therefore ML not only provides a more accurate estimate but also is much lower computational complexity. In order for Newton's algorithm to converge, independently of the position, the error surface defined by (49) and (52) must have only one local minimum. This is true for the case of a ramp edge and can be examined for an arbitrary edge shape by plotting

$$f(\theta)\big|_{x = s_{\hat{\theta}}} \tag{53}$$

in the region $\theta \in [\hat{\theta} - 2L_S, \hat{\theta} + 2L_S]$. If the error surface has several peaks than the Newton's algorithm will have to be repeated over a coarse grid of initial conditions.

6.0 Summary

The CRB is a minimum variance bound on an unbiased estimator. We applied the CRB to edge estimation of an edge with unknown amplitude and position from discrete samples and found that the minimum variance in the edge position estimate is dependent on the sampling period, the SNR, the power in the derivative of the samples, and the angle α between the edge and its derivative. The CRB is therefore generally dependent on sampling position with respect to the edge. However, we showed that provided the sampling rate was above the Nyquist rate of the edge, the discrete sums in the CRB can be replaced by integrals over the continuous-time edge shape and its derivative, which allows us to describe the CRB without reference to the sampling position.

The angle α moves monotonically towards $\pi/2$ as the edge moves away from the end of the observation window on the upside of the edge, so that the CRB becomes constant and minimum when the edge is isolated in the middle of an infinitely large observation window. For sampling above the Nyquist rate the effect of the edge shape on the CRB of an isolated edge is constant regardless of the sampling rate or position (equation (41)). A similar result has only been shown for continuous-time estimators of a Gaussian edge (see [7]) which does not commonly occur in practice.

Practical edge position estimation algorithms use a finite window size to estimate the edge position. We derived a bound on the increase in the CRB due to finite window size for any edge shape (equation (45)) and noted that the fractional increase in the CRB due to finite window size can be kept below a factor of $1/p$ if the observation window extended by p times the edge width. These results allow designers of edge position estimators to intelligently choose the observation window size to keep the computation costs down while ensuring minimal effect of the finite window size on the estimation accuracy.

We showed by simulation that the CRB is a tight bound on the Maximum Likelihood (ML) estimator for both the known and unknown amplitude cases. The ML estimator presented therefore comes very close to the minimum possible estimation error. Despite this level of accuracy, its computational complexity is very low. The known amplitude estimator was two orders of magnitude faster than standard correlation in the example presented.

Appendix A- A Modified Newton's Method for Unknown Amplitude

The cost function to be minimized for the ML estimation is from (48)

$$f(\theta) = \text{II } r\|^2 \tag{54}$$

where

$$r = \left(I - \frac{s_\theta s_\theta^*}{\|s_\theta\|^2}\right)x. \tag{55}$$

Newton's method [123 is an iterative procedure that produces a sequence of estimates $\{\hat{\theta}_1, \hat{\theta}_2, \ldots\}$ that converges quadratically to θ. The iteration is

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \frac{f'(\hat{\theta}_k)}{f''(\hat{\theta}_k)}. \tag{56}$$

First we make some definitions.

$$A = I - \frac{s_\theta s_\theta^*}{\|s_\theta\|^2} \tag{57}$$

and

$$b = \frac{s_\theta}{\|s_\theta\|^2}, \tag{58}$$

Note that $Ah = 0$ and $AA = A$. The first derivative of A is, after some algebraic work

$$A' = -Ad_\theta b^* - (Ad_\theta b^*)^* \tag{59}$$

so that

$$\begin{aligned} f'(\theta) &= -2x^* AA'x \\ &= -2x^* Ad_\theta b^* x \end{aligned} \tag{60}$$

An approximation to the second derivative $f''(\theta)$ is obtained by assuming that r is small, so that all terms containing r can be deleted. Therefore

$$
\begin{aligned}
f'(\theta) &\cong 2r'^{*}r' \\
&= 2\boldsymbol{x}^{*}A'A'\boldsymbol{x} \\
&= 2\,(r^{*}\boldsymbol{d}_{\theta}b^{*}b\boldsymbol{d}_{\theta}^{*}r + \boldsymbol{x}^{*}b\boldsymbol{d}_{\theta}^{*}A\boldsymbol{d}_{\theta}b\boldsymbol{x}) \\
&\cong 2\boldsymbol{x}^{*}b\boldsymbol{d}_{\theta}^{*}A\boldsymbol{d}_{\theta}b\boldsymbol{x}.
\end{aligned}
\tag{61}
$$

Rewriting (59) and (60) without $\boldsymbol{A}$ and $\boldsymbol{b}$ gives

$$
-\frac{1}{2}f'(\theta) = \frac{(\boldsymbol{x}^{*}\boldsymbol{s}_{\theta})}{\|\boldsymbol{s}_{\theta}\|^{2}}\left((\boldsymbol{x}^{*}\boldsymbol{d}_{\theta}) - \frac{(\boldsymbol{x}^{*}\boldsymbol{s}_{\theta})\,(\boldsymbol{d}_{\theta}^{*}\boldsymbol{s}_{\theta})}{\|\boldsymbol{s}_{\theta}\|^{2}}\right) \tag{62}
$$

and

$$
\frac{1}{2}f'(\theta) = \left(\frac{(\boldsymbol{x}^{*}\boldsymbol{s}_{\theta})}{\|\boldsymbol{s}_{\theta}\|^{2}}\right)^{2}\left(\|\boldsymbol{d}_{\theta}\|^{2} - \frac{(\boldsymbol{d}_{\theta}^{*}\boldsymbol{s}_{\theta})^{2}}{\|\boldsymbol{s}_{\theta}\|^{2}}\right). \tag{63}
$$

The complexity of this algorithm is not as great as it may first appear. In a digital system the algorithm will search over a fine grid of points rather than a continuous line. The grid spacing will be the resolution desired. There is no point in making the grid spacing smaller than about the square root of the CRB as the system cannot achieve this resolution anyway. Suppose there are K grid points per sample. It is only necessary to store K values of $\|\boldsymbol{d}_{\theta}\|^{2}$, $\boldsymbol{d}_{\theta}^{*}\boldsymbol{s}_{\theta}$ and $\|\tilde{\boldsymbol{s}}_{\theta}\|^{2}$ (defined in Section 5.0) as these numbers will repeat with period of one sample spacing. The power in the edge $\|\boldsymbol{s}_{\theta}\|^{2}$ is then easily found from (33). Because the derivative has $2L_S/T$ nonzero values, the computational complexity of the inner product between the derivative and the data is not a function of the size of the observation window but only requires $2L_S/T$ multiply-accumulates. Because a portion of the vector $\boldsymbol{s}_{\theta}$ remains constant at 1 recalculation of the inner product is not necessary in this region. A portion of $\boldsymbol{s}_{\theta}$ is zero and calculation of the inner product in this region is unnecessary. If the vector is divided up into three regions, zero, one and the edge, then the inner product in the zero region is zero, the inner product in the one region changes by the sum of the samples between $\hat{\theta}_{k+1}$ and $\hat{\theta}_{k}$, and the inner product in the edge region requires $2L_S/T$ multiply-accumulates. All required values for $\tilde{\boldsymbol{s}}_{\theta}$ and $\boldsymbol{d}_{\theta}$ can be stored in K vectors of length $2L_S/T$. Therefore, the total number of operations required for one iteration is $4L_S/T$ multiply-accumulates and A additions for the inner products, where A is the integer truncation of $\hat{\theta}_{k+1} - \hat{\theta}_{k}$, and seven multiplies and three

additions are required for the update. After the first few iterations the algorithm will get within a few samples of the answer and the value of A will be small from then on so that the computational complexity per iteration will be dominated by the $4L_S/T$ multiply-accumulates.

Appendix B- A Modified Newtons Method for Known Amplitude

The cost function to be minimized for the ML estimator with known amplitude is from (52)

$$f(\theta) = \boldsymbol{x}^* \boldsymbol{s}_\theta - \frac{a}{2}\|\boldsymbol{s}_\theta\|^2. \tag{64}$$

Newton's method [12] is an iterative procedure described in Appendix A. It remains to find the first and second derivatives of (63). This much simpler than for the unknown amplitude case. It is simple to check that

$$f'(\theta) = (\boldsymbol{x} - a\boldsymbol{s}_\theta)^* \boldsymbol{d}_\theta. \tag{65}$$

An approximation to the second derivative is obtained by assuming that $\boldsymbol{x} - a\boldsymbol{s}_\theta$ is small so that all terms containing it can be deleted. Therefore

$$f''(\theta) \cong -2a\|\boldsymbol{d}_\theta\|^2, \tag{66}$$

so that

$$-\frac{f'(\theta)}{f''(\theta)} = \boldsymbol{x}^* \frac{\boldsymbol{d}_\theta}{a\|\boldsymbol{d}_\theta\|^2} - \frac{(\boldsymbol{d}_\theta^* \boldsymbol{s}_\theta)}{\|\boldsymbol{d}_\theta\|^2}. \tag{67}$$

The iteration can be performed to a resolution of T/K by storing the K scalars associated with the second term on the right-hand side of (66) and the K length $2L_S/T$ vectors associated with the inner product in the first term on the right-hand side of (66). The calculation therefore requires only $2L_S/T$ multiply accumulate operations to perform the inner product in the first term on the right-hand side of (66) and then two additions to complete the update.

References

[1] R. E. Hughlett and K. A. Cooper, "A Video Based Alignment System for X-Ray Lithography", Journal of Vacuum Science and Technology, B 9 (6), Nov/Dec 1991.

[2] S. Mallat and S. Zhong, "Characterization of Signals from Multiscale Edges", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 7, July 1992.

[3] H. H. Baker and T. 0. Binford, "Depth from Edge and Intensity Based Stereo", in Proc. 7th Int. Joint Conf. on Artificial Intelligence, 1979, page 631.

[4] P. Bouthemy, "A Maximum Likelihood Framework for Determining Moving Edges", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, May 1989.

[5] H. L. Van Trees, *Detection, Estimation and Modulation Theory,* Part 1, Wiley, 1968.

[6] A. Gatherer and T. H.-Y. Meng, "A Lower Bound on alignment Accuracy and Subpixel resolution in Lithography", Journal of Vacuum Science and Technology, B 9 (6), Nov/Dec 1991.

[7] R. Kakarala and A. 0. Hero, "On Achievable Accuracy in Edge Localization", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 7, July 1992.

[8] J. Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, July 1986.

[9] P. D. Hyde and L. S. Davis "Subpixel Edge Estimation", Pattern Recognition, Vol. 16, No. 4, 1983.

[10] R. *N.* Bracewell, *The Fourier Transform and Its Applications,* 2nd Edition, McGraw-Hill, 1965, page 122.

[11] A. A. Giordano and F. M. Hsu, *Least Square Estimation with Applications to Digital Signal Processing,* Wiley, 1985.

[12] D. G. Luenberger, *Linear and Nonlinear Programming,* 2nd Edition, Addison-Wesley, 1989, Chapter 7.

[13] M. Viberg, B. Ottersten and T. Kailath, "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting", IEEE Transactions on Signal Processing, Vol. 39, No. 11, November 1991.