ESTIMATION OF LAG IN MISREGISTERED, AVERAGED IMAGES

R.J. Carroll,[1] Peter Hall,[2] D. Ruppert[3]

ABSTRACT

In problems where the recorded signal comprises two or more bands (i.e. wavelengths), it is often the case that the bands are not perfectly aligned in the focal plane. Such misregistration may be caused by atmospheric or oceanographic effects, which refract different bands to differing degrees; or it may result from features of the design of the recording equipment. This paper develops two methods for estimating the lag, or amount by which the bands are out of alignment on a pixel grid, in cases where the recorded data are obtained by pixel averaging. Our two techniques are applicable directly to the signal domain and are based on penalized least squares (or errors–in–variables) and maximum cross–covariance, respectively. They employ mathematical interpolation to accommodate the effect of pixel averaging. We introduce a concise and tractable asymptotic model for comparing the performance of the two techniques. The model predicts that the techniques should perform similarly. This is borne out by simulation studies and by applications to real data.

KEY WORDS AND PHRASES. Cross–covariance, errors–in–variables, image analysis, interpolation, lag, penalized least squares, pixel, signal-domain.

SHORT TITLE. Misregistered Averaged Images.

AMS SUBJECT CLASSIFICATION. Primary 62G05, 94A34; Secondary 62J99.

[1] Department of Statistics, Texas A&M University, College Station, TX 77843, USA.

[2] Center for Mathematics and its Applications, Australian National University, Canberra, ACT 2601, and CSIRO Division of Mathematics and Statistics, Sydney, AUSTRALIA.

[3] School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA.

1. INTRODUCTION

Most imaging devices record data in at least two different wavelengths, called bands. For example, a camera recording visible light might register just three different bands, corresponding to the three primary colours. Certain sorts of devices do such recording explicitly, using separate recorders each fitted with narrow–pass filters. Transparent materials such as air, water and glass refract the bands to differing degrees, so that a point source in the image plane (i.e. the plane of the object) is recorded as multiple or almost–coalesced points in the focal plane (i.e. the plane at which the camera records the image). This phenomenon is often known as misregistration. It may be considered to be the result of misalignment of the bands at the focal plane, although of course the manner by means of which the misregistration occurs need not be the result of a physical misalignment inside the recording device. Chromatic aberration of a lens is a well known, albeit relatively complex, example of this phenomenon in the context of imaging at the earth's surface. Simpler examples that are very important in the case of satellite imaging, or imaging from a high–flying aircraft, include band–dependent refraction of electromagnetic radiation by atmospheric effects or water.

For example, certain bands of electromagnetic radiation pass through relatively deep water with only moderate absorption, and so may be used to investigate underwater features. However, water refracts those bands in such a way that they are recorded by a camera in a satellite or aircraft as though they emanated from a position which does not quite correspond to the precise latitude and longitude of their source. Therefore, these bands will be recorded out of register with other wavelengths that emanate from the surface of the water. This can be critical if the precise location of undersea features is to be determined by comparing their positions relative to surface features, like coastlines.

Thus, even in the absence of systematic errors contributed by the manner of construction of recording devices, significant band–to–band misregistration errors may occur. These may be compounded by the imaging equipment. For example, different bands might be recorded in slightly different positions, either because different lenses or pixel arrays are used to record those bands, or because the same lens or array is used but at slightly different times, on a moving camera platform.

This kind of misregistration may occur using either analogue or digital imaging equipment. When the image is recorded digitally, the misregistered bands are averaged over independent picture elements, or pixels. These two forms of image degradation may be compounded by other problems, which might be conveniently included into a noise component of an image model. Our purpose in the present paper is to suggest two different signal–domain methods for correcting band–to–band misregistration, in the presence of pixel averaging and also, possibly, less systematic forms of degradation.

Following standard terminology we shall use the term "signal" to denote the vector of different bands of digitized data. A more concise, mathematical definition of the signal will be provided in Section 2. For the sake of simplicity we shall concentrate on the case of just two bands, although it will be clear that our techniques have straightforward application to higher dimensional circumstances. We shall use the term "lag" to denote the amount by which the two bands are out of register. Typically, lag is measured in terms of a fraction of a pixel width. For example, specifications for some Landsat imaging equipment call for misregistration to be limited to 0.2 pixels between bands in the same focal plane (Berman *et al.* 1990, p. 987).

We shall introduce tractable and concise asymptotic models for analysing our two methods. These models demonstrate that certain specific conditions must be met if the lag is to be estimated consistently. For example, in the case of data involving n pixels and a pixel width h, our model predicts that both h and $(nh^4)^{-1}$ must be small (i.e., converge to zero, in strict asymptotic terms) if the lag is to be estimated consistently, assuming that lag is expressed as a proportion of pixel width. Again, these limitations are borne out by our numerical work; see Sections 4 and 5 for the latter.

As indicated above, one of the strengths of our work is the way in which qualitative conclusions based on intuition or on specific examples of data analysis are given quantitative meaning and accorded general significance using theoretical argument. Our theory is admittedly a little complex, because the very natures of band–to–band misregistration and pixel averaging produce an intrinsically difficult problem. Nevertheless, our theoretical analysis is able to produce substantial information about the sorts of circumstances that are essential to produce good resolution. We show that pixel width must be small, and simultaneously, the value of $r =$ (square root number of pixels) $\times$ (square of pixel width) must be large. Asymptotically, pixel width must tend to zero and r must diverge to infinity if consistency of estimation is to be achieved. Our theory also allows us to quantify the effect on resolution of image smoothness, and to describe the way in which our techniques should be modified to make best use of that smoothness. It shows why a penalty term is necessary in the least squares method, and why maximum cross–covariance does not need a penalty.

Our first technique involves estimating lag by using penalized least squares. We suggest interpolation methods for accommodating the effect of pixel averaging and employ penalized least squares as a goodness–of–fit criterion to effect estimation. In this formulation the squared error represents the difference between the recorded signal in one band and an estimate of that signal computed from another band. The penalty term is necessary to counteract problems caused by correlation arising from both the natural dependence structure of noise in the recorded signal and from additional correlations introduced during signal estimation. The method of penalized least squares may also be interpreted as an errors–in–variables method.

Our second technique involves estimating, and then maximizing, the cross–covariance from one signal and its estimate. This method is conceptually more elegant than the first in that it does not require the treatment of any penalty terms. However, the techniques employ the same methods of interpolation and signal estimation. These aspects are dealt with in depth in Section 2, which also develops methodology for the penalized least squares estimator. Section 3 introduces the method of maximum cross–covariance, Section 4 summarizes simulation studies, and in Section 5 we apply the techniques to real data.

Signal analysis problems involving misregistered but nonaveraged signals date back at least to Hamon & Hannan (1974). See also Pham *et al.* (1987) and Hannan & Thomson (1988). These authors tackle only the one dimensional case. In problems of image analysis, the traditional way of aligning misregistered bands is to identify areas of rapid change in the signals, such as edges, and to concentrate on aligning these. See for example Anuta *et al.* (1984), Wrigley *et al.* (1985) and Tian & Huhns (1986). However, this approach ignores a great deal of information contained in relatively low–frequency parts of the signal; see Berman *et al.* (1990). In contrast, the approach developed in the present paper is tailored to applications to relatively smooth contexts, for example, to images without discontinuities.

We should also mention work in the related area of image deformation registration. Here, two images may be misregistered in a nonuniform way. In particular, two signals may be perfectly matched in certain regions, but may be lagged in others. See for example Amit, Grenander and Piccioni (1991), Bajcsy and Kovacic (1989) and Bookstein (1989).

It is possible to construct Fourier–based versions of the techniques described in this paper. One of these uses interpolation methods to reconstruct the original signal plus an enhancement of the noise from the pixel–averaged signal and then minimizes the difference between two estimates of the Fourier transforms of the original lagged signals in the continuum, noting that, if two functions differ only in terms of a lag, θ, then their Fourier transforms differ only by the multiplicative factor $e^{it\theta}$. Alternatively, one may work directly with discrete Fourier transforms, using interpolation to accommodate the effect of pixel averaging on the amount of the lag. However, neither of these approaches has a particular theoretical advantage over signal-domain methods from the viewpoint of convergence rates when assessed by using our theoretical model. Furthermore, the computational labors of signal-domain and Fourier–domain methods are comparable. We have chosen to discuss signal-domain methods because of their intrinsic statistical as well as engineering interest. We were motivated in part by a desire to explain the poor performance in a practical image analysis application of a non–penalized version of the technique developed in Section 2.

2. ESTIMATION BY PENALIZED LEAST SQUARES

2.1. Summary

In Section 2.2 we describe a model for a degraded two–band signal (X_{1i}, X_{2i}). Here, i denotes a d–vector pixel index with integer elements, and the band X_{ji} for $j = 1, 2$ represents an average of the jth band of the true signal over the ith pixel, plus noise. We take pixel width h to be small, and the number of pixels n to be large; in asymptotic terms, this means that $h = h(n) \to 0$ as $n \to \infty$. It is supposed that the two bands are out of alignment by an amount θh and it is desired to estimate θ.

Section 2.3 describes methods of interpolation that form the basis for our methodology. Of course, there are ways of interpolating other than those we discuss. Some techniques based on Fourier analysis are similar to our interpolation methods with $r = 2$, in the notation of Section 3, where, in effect, r denotes the number of derivatives required of the underlying signal. The interpolation methods described in Section 2.3 are intended only as examples and are not exhaustive.

Section 2.4 introduces the estimation procedure, which is equivalent to penalized least squares with a negative penalty or a nonlinear errors–in–variables method. An asymptotic theory for the estimator is outlined. In particular, we show that, as $h \to 0$ and $n \to \infty$, the bias of the estimator is of order h^{r-1} and the variance is of order $(nh^4)^{-1}$.

2.2. Model

We represent the signal by a "common part", which is recorded similarly in each band, and a "differing part", which is recorded differently in the bands. The "common part" is taken to be the average of a function f over respective pixels, albeit lagged by the amount θh. The "differing part" of the signal is represented by a random variable ϵ; see (2.1) and (2.2) below. In effect, certain features (the "common part" of the image) may be matched by re–alignment, but other features (the "differing part") cannot be matched so easily.

Suppose that the signal is in d dimensions, with $d = 1$ representing classical problems in signal analysis and $d = 2$ modeling image analysis problems. Let f denote a function from $\mathbb{R}^d$ to $\mathbb{R}$. The argument of f represents position in the image plane; the value of f at the position, x say, represents the intensity of the "common part" of the signal at x; and the corresponding intensity of the signal in the film plane is obtained by averaging over pixels. If the pixels are of width h in each dimension and if there is a lag $\theta^{(l)}$ in the lth dimension ($1 \le l \le d$), then, in the case of pixel i, the part of the signal common to both bands and averaged over pixels is given by

$$I_i(\theta) = h^{-d} \int_{\substack{i^{(l)}h+\theta^{(l)}h \le x^{(l)} \le (i^{(l)}+1)h+\theta^{(l)}h, \\ 1 \le l \le d}} f(x)\,dx \quad \text{for} \quad i \in \boldsymbol{Z}^d\,, \tag{2.1}$$

where superscripts index vector elements. Put simply, this represents the average of f over the cube of volume h^d represented by pixel i.

Of course, the lag represented by θ could be incorporated into the function f; if we were to replace f by $f(\cdot + \theta h)$, then $I_i(\theta)$ could be replaced by $I_i(0)$. However, as explained in the Introduction our aim is to model the effect of two bands being lagged by *different* amounts, or equivalently, being out of alignment.

That part of the recorded signal in the first band is taken to be

$$X_{1i} = I_i(0) + \epsilon_{1i} \quad \text{for} \quad i \in \mathcal{I} \subseteq \boldsymbol{Z}^d\,, \tag{2.2a}$$

and that in the second band to be

$$X_{2i} = I_i(\theta_0) + \epsilon_{2i} \quad \text{for} \quad i \in \mathcal{I} \subseteq \boldsymbol{Z}^d\,, \tag{2.2b}$$

where $\{\epsilon_{1i}\}$ and $\{\epsilon_{2i}\}$ represent stochastically independent, second–order stationary processes with zero means, and θ_0 represents the time lag. Our aim is to estimate θ_0. The index set $\mathcal{I}$ represents the range of pixels over which data are recorded.

The model (2.2) for misaligned bands is similar to one proposed by Hamon & Hannan (1974), although those authors did not consider the issue of pixel averaging. By regarding f as a fixed function, rather than a random process, we are in effect conditioning on the common part of the misaligned bands. The methodology and properties that we shall relate in the case of fixed f are of course available when f is random. By confining attention to fixed f we are actually treating a more general result, since convergence rates for random f may be obtained directly from those for fixed f, and may thereby be seen to be identical.

The processes $\{\epsilon_{1i}\}$ and $\{\epsilon_{2i}\}$ may themselves be derived by averaging. For example it may happen that, taking $I_i(\theta, f)$ to denote the right hand side of (2.1), we have $\epsilon_{1i} = I_i(0, \xi_1)$ and $\epsilon_{2i} = I_i(\theta_0, \xi_2)$ for random functions ξ_1 and ξ_2. Thus, the two bands may be regarded as averaged and lagged interpolations of processes which are similar in that they share a component f, upon which we condition. The other components of each band are such that they are independent and stationary, conditional on the common component.

In practice the ϵ processes will often be small relative to the common part of the signals. That is, the ratio of the mean square of ϵ to that of I would be small. Our assumption that the ϵ processes have zero mean is stronger than absolutely necessary; we require only common mean, and even that may be dropped if the signal components X_{1i} and X_{2i} are centered prior to analysis. This change requires only minor modifications to the asymptotic theory developed in Section 2.5. In particular, the convergence rates are unaltered.

One of our methods will require us to estimate the covariance function of the ϵ_j–processes, $\rho_j(k) = E(\epsilon_{ji}\, \epsilon_{j,i+k})$. For this purpose it is convenient to assume a multivariate form of m–correlation, such as

$$\rho_j(k) = 0 \quad \text{if} \quad |k^{(l)}| > m\,, \quad \text{for some} \quad l, \quad 1 \le l \le d, \tag{2.3}$$

where $k = (k^{(1)}, \ldots, k^{(d)})$. The case where the ϵ_1 and ϵ_2 processes are dependent may be accommodated by the penalized least squares method provided we estimate the associated covariance function. However, the main features of our techniques are much clearer if we treat the case where ϵ_1 and ϵ_2 are independent.

2.3. Interpolation

Our methodology is based on approximating the signal component X_{2i} by an interpolated version of the component X_{1i}, the interpolation depending on the unknown θ. We estimate θ_0 by choosing θ to optimize the approximation. It is convenient to treat separately the cases of signals in $d = 1$ and $d = 2$ dimensions.

2.3.1 *Interpolation in one dimension.* We begin by describing interpolation approximations to derivatives. Let g be a univariate function with at least r bounded, continuous derivatives. Given an integer $0 \le j \le r - 1$, let $\mathcal{C}_{jr} = \{c_{jrk}, -\infty < k < \infty\}$ denote a sequence of constants, all but a finite number of them zero, such that for any function g as described above,

$$h^{-j} \sum_k c_{jrk}\, g(x + kh) = g^{(j)}(x) + c_{jr}\, h^{r-j}\, g^{(r)}(x) + o(h^{r-j}) \tag{2.4}$$

as $h \to 0$. The constant c_{jr} depends on the collection $\mathcal{C}_{jr}$, not on g.

The reader is referred to Steffensen (1950, Section 7) and Abramowitz & Stegun (1965, Chapter 25) for a detailed account of different forms of interpolation for numerical differentiation. There are two main approaches: one sided interpolation, where (for interpolation on the right hand side) $c_{jrk} = 0$ for $k < 0$; and two sided interpolation, where c_{jrk} is nonzero for both positive and negative values of k.

Examples of one sided interpolation include Newton's formula, which may be written in the form

$$h^{-j} \sum_{p=j}^{r-1} \frac{1}{p!}\, C_{jp}\, \Delta^p\, g(x) = g^{(j)}(x) - c_{jr}\, h^{r-j}\, g^{(r)}(x) + o(h^{r-j})\,, \tag{2.5}$$

for some constants c_{jr},where the constants C_{jp} are given by (2.8) below, and $\Delta g(x) = g(x + h) - g(x)$, with successive iterations being defined in the obvious way. In this notation, c_{jrk} in (2.4) equals the coefficient of $g(x + kh)$ on the left hand side of (2.4). For some constants d_{jr} one form of two sided interpolation is given by

$$h^{-j} \sum_{p=\frac{1}{2}j}^{r-1} \frac{C'_{j,2p}}{(2p)!}\, \delta^{2p}\, g(x) = g^{(j)}(x) - d_{jr} h^{2r-j} g^{(2r)}(x) + o(h^{2r-j}) \tag{2.6}$$

when j is even, and by

$$h^{-j} \sum_{p=\frac{1}{2}(j+1)}^{r-1} \frac{C'_{j+1,2p}}{(2p-1)!(j+1)}\, \Box\delta^{2p-1}\, g(x) = g^{(j)}(x) - d_{jr} h^{2r-j-1} g^{(2r-1)}(x) + o(h^{2r-j-1}) \tag{2.7}$$

when j is odd, where r has been replaced by $2r$ in the first formula and by $2r-1$ in the second, and where $\delta g(x) = g(x+h/2) - g(x-h/2)$, $\Box g(x) = \frac{1}{2}\{g(x+h/2)+g(x-h/2)\}$, and successive iterations are defined in the obvious way. The constants C_{jp}, C'_{jp} are defined by

$$\begin{aligned} x(x-1)\ldots(x-p+1) &= \sum_{j=0}^{p} x^j \, \frac{C_{jp}}{j!}\,, \\ x\left(x+\tfrac{1}{2}\,p-1\right)\left(x+\tfrac{1}{2}\,p-2\right)\ldots\left(x-\tfrac{1}{2}\,p+1\right) &= \sum_{j=0}^{p} x^j \, \frac{C'_{jp}}{j!}\,, \end{aligned} \tag{2.8}$$

respectively. Formulae (2.5)–(2.8) are taken from Steffensen (1950, Section 7) with minor modifications. Special cases are given in Table 1.

Next we use these interpolation formulae to develop an approximation to X_{2i}. Given constants c_{jrk} with the property (2.4), define

$$\widehat{X}_{2i}(\theta) = X_{1i} + \sum_{j=1}^{r-1} \left(\theta^j/j!\right) \sum_k c_{jrk}\, X_{1,i+k} = I_{ir}(\theta) + \eta_{ir}(\theta)\,, \tag{2.9}$$

where

$$I_{ir}(\theta) = I_i(0) + \sum_{j=1}^{r-1} \left(\theta^j/j!\right) \sum_k c_{jrk}\, I_{i+k}(0)\,, \tag{2.10}$$

$$\eta_{ir}(\theta) = \epsilon_{i1} + \sum_{j=1}^{r-1} \left(\theta^j/j!\right) \sum_k c_{jrk}\, \epsilon_{1,i+k}\,. \tag{2.11}$$

The random variable η_{ir} has zero mean and $I_{ir}(\theta)$ represents an approximation to $I_i(\theta)$ with error of size h^r. To appreciate this point observe that with $F(x) = \int^x f(u)\,du$ we have

$$\begin{aligned} I_i(\theta) &= h^{-1}\left[F\{(i+1)h+\theta h\} - F(ih+\theta h)\right] \\ &= I_i(0) + \sum_{j=1}^{r-1} \frac{(\theta h)^j}{j!}\, h^{-1}\left[F^{(j)}\{(i+1)h\} - F^{(j)}(ih)\right] + \frac{(\theta h)^r}{r!}\, f^{(r)}(ih) + o(h^r) \\ &= I_{ir}(\theta) - \sum_{j=1}^{r-1} \frac{(\theta h)^j}{j!}\, h^{-1}\Big(\Big[h^{-j} \sum_k c_{jrk}\, F\{(i+1+k)h\} - F^{(j)}\{(i+1)h\}\Big] \\ &\qquad - \Big[h^{-j} \sum_k c_{jrk}\, F\{(i+k)h\} - F^{(j)}(ih)\Big]\Big) + \frac{(\theta h)^r}{r!}\, f^{(r)}(ij) + o(h^r) \\ &= I_{ir}(\theta) - h^r\, J_{ir}(\theta)\,, \end{aligned} \tag{2.12}$$

where

$$J_{ir}(\theta) = f^{(r)}(ih)\left(\sum_{j=1}^{r-1} \frac{\theta^j}{j!} c_{jr} - \frac{\theta^r}{r!}\right) + o(1)\,.$$

In this notation, $\widehat{X}_{2i}(\theta) = I_i(\theta) + h^r\, J_{ir}(\theta) + \eta_{ir}(\theta)$. As we show in Section 2.4, we estimate θ_0 by comparing $\widehat{X}_{2i}(\theta)$ with $X_{2i} = I_i(\theta_0) + \epsilon_{2i}$.

2.3.2 *Interpolation in two dimensions.* Let $\mathcal{C}_{jr} = \{c_{jrk}, -\infty < k < \infty\}$ denote the sequence introduced in Section 2.3.1, enjoying property (2.4). Now let g be a bivariate function with r bounded, continuous derivatives of all types. Then for $j_1 + j_2 < r$ an approximation to $g^{(j_1,j_2)}(x_1, x_2)$ is given by

$$\begin{aligned} g_r^{(j_1,j_2)}(x_1,x_2) &= h^{-(j_1+j_2)} \sum_{k_1} \sum_{k_2} c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, g(x_1+k_1h, x_2+k_2h) \\ &= g^{(j_1,j_2)}(x_1,x_2) + h^{r-j_1-j_2} \big\{ c_{j_1,r-j_2}\, g^{(r-j_2,j_2)}(x_1,x_2) \\ &\qquad + c_{j_2,r-j_1}\, g^{(j_2,r-j_2)}(x_1,x_2) \big\} + o\big(h^{r-j_1-j_2}\big)\,. \end{aligned}$$

By analogy with (2.9) our approximation to X_{2i} is

$$\begin{aligned} \widehat{X}_{2i}\big(\theta^{(1)},\theta^{(2)}\big) &= X_{1i} + \sum\sum_{1\le j_1+j_2\le r-1} \frac{(\theta^{(1)})^{j_1}\,(\theta^{(2)})^{j_2}}{j_1!\,j_2!} \sum_{k_1}\sum_{k_2} c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, X_{1,i+k} \\ &= I_{ir}\big(\theta_1^{(1)},\theta_2^{(2)}\big) + \eta_{ir}\big(\theta_1^{(1)},\theta_2^{(2)}\big)\,, \end{aligned}$$

where

$$\begin{aligned} I_{ir}\big(\theta^{(1)},\theta^{(2)}\big) &= I_i(0,0) + \sum\sum_{1\le j_1+j_2\le r-1} \frac{(\theta^{(1)})^{j_1}\,(\theta^{(2)})^{j_2}}{j_1!\,j_2!} \\ &\qquad \times \sum_{k_1}\sum_{k_2} c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, I_{i+k}(0,0)\,, \\ \eta_{ir}\big(\theta^{(1)},\theta^{(2)}\big) &= \epsilon_{1i} + \sum\sum_{1\le j_1+j_2\le r-1} \frac{(\theta^{(1)})^{j_1}\,(\theta^{(2)})^{j_2}}{j_1!\,j_2!} \\ &\qquad \times \sum_{k_1}\sum_{k_2} c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, \epsilon_{1,i+k}\,; \end{aligned}$$

compare (2.10) and (2.11).

To appreciate that I_{ir} approximates I_i, first define $F(x,y) = \int^x \int^y f(u,v)\, du\, dv$. The analogue of (2.12) is

$$\begin{aligned} I_i\big(\theta^{(1)},\theta^{(2)}\big) &= h^{-2}\big[F\big\{(i_1+1)h + \theta^{(1)}h, (i_2+1)h + \theta^{(2)}h\big\} \\ &\qquad - F\big\{i_1h + \theta^{(1)}h, (i_2+1)h + \theta^{(2)}h\big\} \\ &\qquad - F\big\{(i_1+1)h + \theta^{(1)}h, i_2h + \theta^{(2)}h\big\} + F(i_1h + \theta^{(1)}h, i_2h + \theta^{(2)}h)\big] \\ &= I_{ir}\big(\theta^{(1)},\theta^{(2)}\big) - h^r\, J_{ir}\big(\theta^{(1)},\theta^{(2)}\big)\,, \end{aligned}$$

where

$$J_{ir}\big(\theta^{(1)},\theta^{(2)}\big) = \sum\sum_{1\le j_1+j_2\le r-1} \frac{(\theta^{(1)})^{j_1}\,(\theta^{(2)})^{j_2}}{j_1!\,j_2!} \times \big\{c_{j_1,r-j_2}\,f^{(r-j_2,j_2)}(i_1h,i_2h) + c_{j_2,r-j_1}\,f^{(j_1,r-j_1)}(i_1h,i_2h)\big\} - \sum\sum_{j_1+j_2=r} \frac{(\theta^{(1)})^{j_1}\,(\theta^{(2)})^{j_2}}{j_1!\,j_2!}\,f^{(j_1,j_2)}(i_1h,i_2h)\,.$$

In this notation, $\widehat{X}_{2i}\big(\theta^{(1)},\theta^{(2)}\big) = I_i\big(\theta^{(1)},\theta^{(2)}\big) + h^r\,J_{ir}\big(\theta^{(1)},\theta^{(2)}\big) + \eta_{ir}\big(\theta^{(1)},\theta^{(2)}\big)$, which represents an approximation to $X_{2i} = I_1\big(\theta_0^{(1)},\theta_0^{(2)}\big) + \epsilon_{2i}$.

2.4. Estimation

Section 2.4.1 motivates our estimator $\widehat{\theta}$ of the unknown θ in terms of minimizing a sum of penalized squared discrepancies between the observed signals and linear approximations. Section 2.4.2 discusses computation of the penalty term and Section 2.4.3 describes asymptotic theory for the estimator. In terms of asymptotics we assume that the pixel width h shrinks to zero as the number of pixels n, i.e., the number of elements of the set $\mathcal{I}$, see (2.2), diverges. It emerges from our analysis that consistency requires $nh^4 \to \infty$ as $n \to \infty$ and $h \to 0$.

2.4.1 *Motivation.* In Section 2.3 we defined an approximation $\widehat{X}_{2i}(\theta)$ to X_{2i}. We now suggest choosing θ to optimize the performance of this approximation. To determine the sense of the "optimum", consider the simple squared distance measure, $S_1(\theta) = \sum_i \{\widehat{X}_{2i}(\theta) - X_{2i}\}^2$, where the sum over values i is the index set $\mathcal{I}$ on which data are recorded; see (2.2). We claim that minimizing $S_1(\theta)$ with respect to θ is inadequate. It produces a biased, inconsistent estimator of θ_0, as we now relate.

Recall from Section 2.3 that $\widehat{X}_{2i}(\theta) = I_i(\theta) + h^r\,J_{ri}(\theta) + \eta_{ri}(\theta)$, and from the model (2.2) that $X_{2i} = I_i(\theta_0) + \epsilon_{2i}$. Therefore,

$$\widehat{X}_{2i}(\theta) - X_{2i} = I_i(\theta) - I_i(\theta_0) + h^r\,J_{ir}(\theta) + \eta_{ir}(\theta) - \epsilon_{2i}\,. \tag{2.13}$$

In the case of $d = 1$ dimensions, as $\theta \to \theta_0$ we have $J_{ir} = J_{ir}(\theta_0) + o(1)$,

$$I_i(\theta) - I_i(\theta_0) = c_i h(\theta-\theta_0) + o(h|\theta-\theta_0|)\,, \tag{2.14}$$

$$\eta_{ir}(\theta) = \eta_{ir}(\theta_0) + h(\theta-\theta_0)\sum_k c_{irk}\,\epsilon_{1,i+k} + o_p(h|\theta-\theta_0|)\,, \tag{2.15}$$

where c_i denotes a constant which, under mild conditions on f, is of order 1. Comparing the right hand sides of (2.14) and (2.15) we see that both $I_i(\theta) - I_1(\theta_0)$ and $\eta_{ir}(\theta)$ contain terms in $h(\theta-\theta_0)$. Thus, noting the form of the right hand side of (2.13), it becomes clear that the estimator of θ which results from simply minimizing $S_1(\theta)$ will not be consistent unless we do something to take account of the dependence of $\eta_{ir}(\theta)$ on θ. The same is true for $d \geq 2$.

It is not difficult to see that the most significant contribution to $S_1(\theta)$ of the term $h(\theta-\theta_0)$ in $\eta_{ir}(\theta)$ comes from the square of $\eta_{ir}(\theta)$. The cross–product terms involving $\eta_{ir}(\theta)$ make a relatively lesser contribution owing to the fact that the errors ϵ_{1i} have zero mean. Furthermore the sum of the squares of the $\eta_{ir}(\theta)$'s is closely approximated by its expected value. Thus, we might first estimate $\tau(\theta) = \sum_i E\{\eta_{ir}(\theta)^2\}$ by $\widehat{\tau}(\theta)$ say, and then choose $\theta = \widehat{\theta}$ to minimize

$$S(\theta) \equiv S_1(\theta) - \widehat{\tau}(\theta) = \sum_i \left\{\widehat{X}_{2i}(\theta) - X_{2i}\right\}^2 - \widehat{\tau}(\theta)\,. \tag{2.16}$$

It is not crucial that $\widehat{\tau}(\theta)$ be consistent for $\tau(\theta)$, but it is essential that $\widehat{\tau}(\theta)$ represent a good approximation to those parts of $\tau(\theta)$ that depend on θ.

By construction this method uses a linear interpolant of X_{1i} to approximate X_{2i}. Of course the problem of estimating θ has inherent symmetry, and we could instead have approximated X_{1i} by an interpolant of X_{2i}, say $\widehat{X}_{1i}(-\theta)$. Arguing thus and letting $\widehat{\tau}_1(\theta) = \tau(\theta)$ and $\widehat{\tau}_2(\theta)$ denote the respective versions of $\tau(\theta)$, we see that we would choose $\theta = \widehat{\theta}$ to minimize

$$\sum_i \left[\left\{\widehat{X}_{2i}(\theta) - X_{2i}\right\}^2 + \left\{\widehat{X}_{1i}(-\theta) - X_{1i}\right\}^2\right] - \widehat{\tau}_1(\theta) - \widehat{\tau}_2(-\theta) \tag{2.17}$$

instead of $\widehat{S}(\theta)$ at (2.16).

This approach uses the available information more effectively and so it comes as no surprise to learn that it reduces mean squared error of the estimator $\widehat{\theta}$. However it does not reduce the order of magnitude of mean squared error, only the constant multiplying that order. Working with the criterion (2.17) is notationally cumbersome, and so we shall develop theory only for the simpler criterion (2.16). The manner of transition from one to the other will be clear.

Our estimation may also be motivated by consideration of an errors–in–variables nonlinear regression model (Fuller, 1987). Let $Y_i = X_{2i} - X_{1i}$, $\epsilon_{i*} = \epsilon_{2i} - \epsilon_{1i}$ and $Z_{ij} = (j!)^{-1} \sum_k c_{jrk}\, I_{i+k}(0)$. Except for terms of order h^r, $Y_i = \sum_{1\leq j\leq r-1} \theta_0^j\, Z_{ij} + \epsilon_{i*}$, which is a nonlinear regression model. Consistent estimates of θ_0 would be obtained by minimizing

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{j=1}^{r-1}\theta^j\,Z_{ij}\right\}^2 .$$

However, Z_{ij} cannot be observed, and instead we estimate it by $\widehat{Z}_{ij}=(j!)^{-1}\sum_k c_{jrk}\,X_{1,i+k}$ $=Z_{ij}+V_{ij}$. Nonlinear regression of Y_i on $\{\widehat{Z}_{ij}\}$ would produce inconsistent estimators of θ because of the errors–in–variables phenomenon; $\widehat{Z}_{ij}$ is an error–prone version of Z_{ij}. Indeed,

$$\sum_{i=1}^{n}\left\{Y_i-\sum_{j=1}^{r-1}\theta^j\,\widehat{Z}_{ij}\right\}^2=\sum_{i=1}^{n}\left\{\widehat{X}_{2i}(\theta)-X_{2i}\right\}^2=\sum_{i=1}^{n}\left\{I_i(\theta)-I_i(\theta_0)+\epsilon_{2i}-\eta_{ir}(\theta)+O(h^r)\right\}^2,$$

has expectation $\sum_{1\le i\le n}\{I_i(\theta)-I_i(\theta_0)\}^2+\tau(\theta)$, up to terms not depending on θ. Since we want to minimize the first term in this sum it makes sense to subtract $\widehat{\tau}(\theta)$ from the observed sum of squares, leading to (2.16).

2.4.2 *Estimation of* $\tau(\theta)$. We treat separately the cases of one and two dimensional signals. In the case $d=1$ observe from (2.11) that

$$E\{\eta_{ir}(\theta)^2\}=\rho(0)+2\sum_{j=1}^{r-1}\sum_{k}\left(\theta^j/j!\right)c_{jrk}\,\rho(k)$$
$$+\sum_{j_1=1}^{r-1}\sum_{j_2=1}^{r-1}\sum_{k_1}\sum_{k_2}\frac{\theta^{j_1+j_2}}{j_1!\,j_2!}\,c_{j_1rk_1}\,c_{j_2rk_2}\,\rho(k_1-k_2)\,,$$

where $\rho(k)=\rho_1(k)=E(\epsilon_{1i}\,\epsilon_{1,i+k})$. The first term on the right hand side, $\rho(0)$, does not depend on θ and so may be ignored. Therefore we take

$$\widehat{\tau}(\theta)=n\left\{2\sum_{j=1}^{r-1}\sum_{k}\left(\theta^j/j!\right)c_{jrk}\,\widehat{\rho}(k)\right.$$
$$\left.+\sum_{j_1=1}^{r-1}\sum_{j_2=1}^{r-1}\sum_{k_1}\sum_{k_2}\frac{\theta^{j_1+j_2}}{j_1!\,j_2!}\,c_{j_1rk_1}\,c_{j_2rk_2}\,\widehat{\rho}(k_1-k_2)\right\},$$

where $\widehat{\rho}(k)$ denotes an estimator of $\rho(k)$.

Next we describe how to estimate $\rho(k)$. We assume the m–correlation model (2.3) under which $\rho(k)$ is known to be zero for $|k|>m$. Let $\{b_{rk},-\infty<k<\infty\}$ and b_r denote real numbers such that for a univariate function g having r continuous derivatives,

$$\sum_k b_{rk}\,g(x+kh)=b_r\,h^r\,g^{(r)}(x)+o(h^r)$$

as $h \to 0$. For example we may take $b_{rk} = c_{r,r+1,k}$ and $b_r = 1$. For integers $q \geq 1$ define

$$Y_{qi} = \sum_k b_{rk}\, X_{1,i+kq} = K_{qi} + \zeta_{qi}$$

where

$$K_{qi} = \sum_k b_{rk}\, I_{i+kq}(0) = b_r\, q^r\, h^r\, f^{(r)}(ih) + o(h^r)\,, \qquad \zeta_{qi} = \sum_k b_{rk}\, \epsilon_{1,i+kq}\,.$$

Put

$$\beta_j = \sum_{\substack{k_1,k_2 \\ |k_1-k_2|=j}}\sum b_{rk_1}\, b_{rk_2}\,. \tag{2.18}$$

In this notation

$$t_{rq} = E\big(\zeta_{rqi}^2\big) = \sum_{j=0}^{m/q} \beta_j\, \rho(jq)\,, \tag{2.19}$$

which is estimated by $\widehat{t}_{rq} = n^{-1} \sum_i Y_{qi}^2 = t_{rq} + O_p\big(n^{-1/2} + h^{2r}\big)$.

Formula (2.19) expresses the estimable quantity t_{rq} as a known linear combination of the unknowns $\rho(k)$. The $m+1$ equations $\sum_{0\leq j\leq m/q} \beta_j\, \widehat{\rho}(jq) = \widehat{t}_{rq}$ for $1 \leq q \leq m+1$, are linearly independent in the $m+1$ unknowns $\widehat{\rho}(0), \ldots, \widehat{\rho}(m)$ and may be solved to yield estimators $\widehat{\rho}(k)$ which satisfy $\widehat{\rho}(k) - \rho(k) = O_p\big(n^{-1/2} + h^{2r}\big)$ for $0 \leq k \leq m$. Indeed the equations may be solved recursively, giving $\rho(0), \rho(m), \rho(m-1), \ldots, \rho(1)$, by arguing as follows:

$$\widehat{\rho}(0) = \beta_0^{-1}\, \widehat{t}_{r,m+1}\,, \qquad \widehat{\rho}(m) = \beta_1^{-1}\, \{\{\widehat{t}_{rm} - \beta_0\, \widehat{\rho}(0)\}\,,$$
$$\widehat{\rho}(m-1) = \beta_1^{-1}\, [\widehat{t}_{r,m-1} - \beta_0\, \widehat{\rho}(0) - \beta_2\, \widehat{\rho}\{2(m-1)\}], \ldots\,,$$
$$\widehat{\rho}(1) = \beta_1^{-1}\{\widehat{t}_{r1} - \beta_0\widehat{\rho}(0) - \beta_2\widehat{\rho}(2) - \ldots - \beta_m\widehat{\rho}(m)\}\,,$$

where we take $\widehat{\rho}(k) = 0$ if $k > m$.

In the case of $d = 2$ dimensions we take

$$\begin{aligned}\widehat{\tau}(\theta) = n\bigg\{2 \sum_{1\leq j_1+j_2\leq r-1}\sum \frac{(\theta^{(1)})^{j_1}\, (\theta^{(2)})^{j_2}}{j_1!\, j_2!} \sum_{k_1}\sum_{k_2} c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, \widehat{\rho}(k_1,k_2) \\ + \sum_{1\leq j_1+j_2\leq r-1}\sum\; \sum_{1\leq j_3+j_4\leq r-1}\sum\; \sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4} \frac{(\theta^{(1)})^{j_1+j_3}\, (\theta^{(2)})^{j_2+j_4}}{j_1!\, j_2!\, j_3!\, j_4!} \\ \times c_{j_1,r-j_2,k_1}\, c_{j_2,r-j_1,k_2}\, c_{j_3,r-j_4,k_3}\, c_{j_4,r-j_3,k_4}\, \widehat{\rho}(k_1-k_3, k_2-k_4)\bigg\}\,.\end{aligned}$$

If the noise process is isotropic so that $\rho(j,k)=\rho(-j,-k)=\rho(-j,k)=\rho(j,-k)$, and if we define m–dependence by $\rho(j,k)=0$ if $|j|>m$ or $|k|>m$, then we use the following procedure. Let $b_{r,k}=c_{r-[r/2],r+1-[r/2],k}$ and $d_{r,k}=c_{[r/2],[r/2]+1,k}$. For $q=(q_1,q_2)$, $i=(i_1,i_2)$ and $k=(k_1,k_2)$ define

$$Y_{qi}=\sum_{k_1}\sum_{k_2} b_{r,k_1} d_{r,k_2} X_{1,i+kq}=\sum_{k_1}\sum_{k_2} b_{r,k_1} d_{r,k_2} \epsilon_{1,i+kq}+O(h^r)=\zeta_{q,i}+O(h^r)\,.$$

If we define $\beta_{j,1}$ by (2.18) and $\beta_{j,2}$ by (2.18) but with b_{rk} replaced by d_{rk} then for $1\le q_1,q_2\le m+1$,

$$t_q=E\left(\zeta_{q,i}^2\right)=\sum_{j_1=0}^{[m/q_1]}\sum_{j_2=0}^{[m/q_2]}\beta_{j_1,1}\beta_{j_2,2}\rho\left(j_1q_1,j_2q_2\right)\,.$$

Defining $\widehat{t}_q=n^{-2}\sum_i Y_{q,i}^2$ we estimate $\rho(\cdot,\cdot)$ by solving $\widehat{t}_q=t_q$. Again the bias is of size h^{2r} and the error–about–the–mean of size $n^{-1/2}$, with these formulae also applying to $\widehat{\tau}(\theta)$.

Estimates of $\rho(\cdot,\cdot)$ for other covariance structures can be found by similar techniques.

2.4.3 *Asymptotic theory.* In this section we show that if the underlying signal f has $r\ge 2$ bounded derivatives and if the noise processes ϵ_{1i} and ϵ_{2i} have the m–correlation property, see (2.3), then $\widehat{\theta}$ has bias of size h^{r-1} and error–about–the–mean of size $n^{-1/2}\,h^{-2}$. We shall treat only the case $d=1$ dimension using (2.16); $d=2$ and (2.17) are similar.

Define $\rho(k)=\rho_1(k)=E(\epsilon_{1i},\epsilon_{1,i+k})$,

$$S_1^{-1}=n^{-1}\sum_i f^{(1)}(ih)^2\,,\qquad S_2=n^{-1}\sum_i f^{(1)}(ih)\,f^{(r)}(ih)\,,$$

$$S_3=\sum_{j=1}^{r-1}\frac{\theta_0^j}{j!}\,c_{jr}-\frac{\theta_0^r}{r!}\,,\qquad S_4=-S_1S_2S_3\,,$$

$$\widehat{\alpha}_1=n^{-1}\sum_i\left(\epsilon_{1i}+\sum_{j=1}^{r-1}\frac{\theta_0^j}{j!}\sum_i c_{jrk}\,\epsilon_{1,i+k}-\epsilon_{2i}\right)\left(\sum_{j=1}^{r-1}\frac{\theta_0^{j-1}}{(j-1)!}\sum_k c_{jrk}\,\epsilon_{1,i+k}\right),$$

$$\widehat{\alpha}=\widehat{\alpha}_1-E(\widehat{\alpha}_1)=O_p\left(n^{-1/2}\right),$$

$$\widehat{\beta}=h^{-2}\,S_1\left[\sum_{j=1}^{r-1}\frac{\theta_0^{j-1}}{(j-1)!}\sum_k c_{jrk}\{\widehat{\rho}(k)-\rho(k)\}\right.$$
$$+\frac{1}{2}\sum_{j_1=1}^{r-1}\sum_{j_2=1}^{r-1}\sum_{k_1}\sum_{k_2}\frac{(j_1+j_2)\theta_0^{j_1+j_2-1}}{j_1!\,j_2!}$$
$$\left.\times\, c_{j_1rk_1}\,c_{j_2rk_2}\,\{\widehat{\rho}(k_1-k_2)-\rho(k_1-k_2)\}\right].$$

Let $\widehat{\rho}(k)$ denote an estimator of $\rho(k)$ with bias of $o(h^{r+1})$ and error–about–the–mean of $O_p\big(n^{-1/2}\big)$. The estimator suggested in the previous section has this property, which means that $\widehat{\beta} = O_p(n^{-1/2}\, h^{-2}) + o(h^{r+1})$. We shall prove that

$$\widehat{\theta} - \theta_0 = -S_1\, h^{-2}\, \widehat{\alpha} + \widehat{\beta} + S_4\, h^{r-1} + o_p\big(n^{-1/2}\, h^{-2} + h^{r-1}\big)\,, \tag{2.20}$$

which establishes the claim in the previous paragraph.

Our starting point is formulae (2.14)–(2.16) from which it follows that

$$\tfrac{1}{2}\, \frac{\partial}{\partial\theta}\, S(\theta) = T_1 + T_2 + T_3 - h^2\, S_1^{-1}\, \widehat{\beta}\,,$$

where as $\theta \to \theta_0$, $h \to 0$ and $n \to \infty$,

$$\begin{aligned}
T_1 &= \sum_i \big\{I_i(\theta) - I_i(\theta_0) + h^r\, J_{ir}(\theta)\big\}\big\{I_i^{(1)}(\theta) + h^r\, J_{ir}^{(1)}(\theta)\big\} \\
&= (\theta - \theta_0) \sum_i \Big\{I_i^{(1)}(\theta_0)\Big\}^2 + h^r \sum_i I_i^{(1)}(\theta_0)\, J_{ir}(\theta_0) + o(nh^2|\theta - \theta_0| + nh^{r+1}) \\
&= (\theta - \theta_0)\, nh^2\, S_1^{-1} + nh^{r+1}\, S_2\, S_3 + o(nh^2|\theta - \theta_0| + nh^{r+1})\,, \\
T_2 &= \sum_i \Big[\big\{I_i(\theta) - I_i(\theta_0) + h^r\, J_{ir}(\theta)\big\}\, \eta_{ir}^{(1)}(\theta) + \big\{I_i^{(1)}(\theta) + h^r\, J_{ir}^{(1)}(\theta)\big\}\{\eta_{ir}(\theta) - \epsilon_{2i}\}\Big] \\
&= O_p(n^{1/2} h)\,, \\
T_3 &= \sum_i \{\eta_{ir}(\theta) - \epsilon_{2i}\}\, \eta_{ir}^{(1)}(\theta) - E\Big\{\sum_i \eta_{ir}(\theta)\, \eta_{ir}^{(1)}(\theta)\Big\} = \widehat{\alpha}\,.
\end{aligned}$$

Combining these results we deduce (2.20).

3. ESTIMATION BY MAXIMUM CROSS–COVARIANCE

3.1. The estimator

This section presents an alternative to the penalized least squares estimator. The alternative estimator has two virtues, (a) it is conceptually and computationally simpler than penalized least squares, and (b) it is consistent for a wide class of error processes including, but not limited to, m–dependence.

Let $\widehat{X}_{2i}(\theta)$ be given by (2.9) or its analogue in the two dimensional case, let

$$T(\theta) = n^{-1} \sum_i X_{2i}\, \widehat{X}_{2i}(\theta)\,, \tag{3.1}$$

and define

$$\widehat{\theta} = \arg\max T(\theta)\,. \tag{3.2}$$

To understand the motivation behind $\widehat{\theta}$ notice that by (2.2b) and (3.1),

$$T(\theta) = n^{-1} \sum_i \{I_i(\theta_0) + \epsilon_{2i}\}\big\{I_i(\theta) + h^r\, J_{ir}(\theta) + \eta_{ir}(\theta)\big\}\,.$$

Since $\{\epsilon_{2i}\}$ and $\{\eta_{ir}(\theta)\}$ are independent processes and using the Cauchy–Schwarz inequality we can expect that

$$T(\theta) \approx n^{-1} \sum_i I_i(\theta_0)\, I_i(\theta) \leq n^{-1} \Big\{ \sum_i I_i^2(\theta) \sum_i I_i^2(\theta_0) \Big\}^{1/2}\,,$$

where the equality holds if and only if $I_i(\theta)$ (as a function of i) is linear in $I_i(\theta_0)$. Unless F changes rapidly near 0 or nh, $\sum_i I_i(\theta)^2$ is approximately independent of θ, and then $\widehat{\theta} \approx \theta_0$.

If X_{2i} is a mean zero process then $T(\theta)$ is the sample cross–covariance at lag zero between X_{2i} and $\widehat{X}_{2i}(\theta)$. In what is to follow we assume that X_{1i} and X_{2i} have been centered to have mean equal to zero, and in practice we shall maximize

$$T(\theta) = n^{-1} \sum_i \Big\{ X_{2i}\widehat{X}_{2i}(\theta) + X_{1i}\widehat{X}_{1i}(-\theta) \Big\}\,. \tag{3.3}$$

This procedure is simple to implement since it does not require an estimator of $\rho(\cdot)$. In contrast, penalized least squares could require extensive programming when $d \geq 2$ and the order of dependence, m, is not small. However a disadvantage of the maximum cross–covariance estimator is that as $h \to 0$ its bias converges to zero more slowly than does the bias of penalized least squares, for the same value of r.

3.2. Asymptotic theory

Consider the case $d = 1$. The estimator $\widehat{\theta}$ satisfies the equation

$$0 = T^{(1)}(\widehat{\theta}) = T^{(1)}(\theta_0) + T''(\theta^*)(\widehat{\theta} - \theta_0)\,, \tag{3.4}$$

for an intermediate point θ^*. Asymptotically $\widehat{\theta}$ behaves like $\tilde{\theta} = \theta_0 - T^{(1)}(\theta_0)/T''(\theta_0)$. We shall study the asymptotic behavior of $\tilde{\theta}$.

Observe that

$$T^{(j)}(\theta_0) = \frac{1}{n} \sum_i \{I_i(\theta_0) + \epsilon_{2i}\}\big\{I_i^{(j)}(\theta_0) + h^r\, J_{ir}^{(j)} + \eta_{ir}^{(j)}(\theta_0)\big\}\,. \tag{3.5}$$

One can show that

$$J_{ir}^{(j)}(\theta_0) = O(1) \quad \text{for} \quad j = 1, 2\,, \tag{3.6}$$

and

$$\operatorname{Var}\big(\eta_{ir}^{(j)}(\theta_0)\big) = O(1) \quad \text{for} \quad j = 1, 2\,. \tag{3.7}$$

We shall show below that

$$n^{-1} \sum_{i=1}^{n} I_i(\theta_0)\, I_i^{(1)}(\theta_0) = O(h^5) + O(n^{-1}) \tag{3.8}$$

and

$$n^{-1} \sum_{i=}^{n} I_i(\theta_0)\, I_i^{(2)}(\theta_0) \sim h^2\,, \tag{3.9}$$

where the notation $a_n \sim b_n$ means that a_n/b_n is bounded away from zero and infinity. By (3.5)–(3.9), $n^{-1}\, T^{(1)}(\theta_0) = O_p\big(n^{-1/2}\big) + O\big(h^{\min(5,r)}\big)$ and $n^{-1}\, T''(\theta_0) \sim h^2$, assuming that $n^{-1/2}\, h^{-2} \to 0$. Therefore

$$\tilde\theta - \theta_0 = O_p\big(n^{-1/2}\, h^{-2}\big) + O\big(h^{\min(3,r-2)}\big)\,. \tag{3.10}$$

The first term on the right hand side of (3.9) is the error–about–the–mean, while the second term is bias. To establish (3.8) define $\phi(\theta) = n^{-1} \sum_{1\le i\le n} I_i(\theta_0)\, I_i(\theta)$. Recall that $I_i(\theta) = h^{-1}[F\{(i+1+\theta)h\} - F\{(i+\theta)h\}]$. To simplify notation take $\theta_0 = 0$; this does not cause a loss in generality. Then

$$\begin{aligned}
\phi^{(1)}(0) &= n^{-1} h \sum_{i=1}^{n} \big(h^{-1}[F\{(i+1)h\} - F(ih)]\, h^{-1}[F^{(1)}\{(i+1)h\} - F^{(1)}(ih)]\big) \\
&= n^{-1} h \sum_{i=1}^{n} \big[F^{(1)}\big\{\big(i+\tfrac12\big)h\big\} + (1/6) F^{(3)}\big\{\big(i+\tfrac12\big)h\big\}h^2 + O(h^4)\big] \\
&\qquad \times \big[F^{(2)}\big\{\big(i+\tfrac12\big)h\big\} + (1/6) F^{(4)}\big\{\big(i+\tfrac12\big)h\big\}h^2 + O(h^4)\big] \\
&= O(h^5) + O(n^{-1})\,,
\end{aligned} \tag{3.11}$$

since for any j_1, j_2 we have

$$n^{-1} \sum F^{(j_1)}\big\{\big(i+\tfrac12\big)h\big\}\, F^{(j_2)}\big\{\big(i+\tfrac12\big)\big\} = (nh)^{-1} \int_0^{nh} F^{(j_1)}(x)\, F^{(j_2)}(x)\, dx + O(h^2), \tag{3.12}$$

and for any j,

$$\int_0^{nh} F^{(j)}(x)\, F^{(j+1)}(x)\, dx = \tfrac12 \left\{F^{(j)}(x)\right\}^2 dx \Big|_0^{nh} = O(1)\,.$$

Note that (3.11) proves (3.8). In several places we have assumed that $F^{(j)}(x) = O(1)$ for $x = 0$ and nh and for $j = 1, 2, 3$. Note that $F^{(j)}(0)$ is fixed since the signal is fixed, but

$F^{(j)}(nh)$ could grow rapidly. Therefore it is necessary to assume that $F^{(j)}$ is bounded away from ∞ for $j \le 3$. Without this assumption (3.10) could fail to hold. Similar calculations show that

$$\phi''(0) = h^2\Big\{(nh)^{-1} \int_0^{nh} F^{(1)}(x)\, F^{(3)}(x)\, dx + O(1)\Big\}$$

$$= h^2\Big\{-(nh)^{-1} \int_0^{nh} \big(F^{(2)}(x)\big)^2\, dx + O(nh)^{-1} + O(1)\Big\} \sim h^2\,, \qquad (3.13)$$

since we shall assume that $\int_{0<x<nh} \big(F^{(2)}(x)\big)^2\, dx \sim nh$ and $n^{-1} = o(h^2)$. We can think of the left hand side of (3.13) as the "signal" or "information about θ" in the data. Notice that if $F^{(2)}(x) \equiv 0$, then $F(x) = F(0) + F^{(1)}(0)\, x$ and $I_i(\theta) = h^{-1}[F\{i+1+\theta)h\} - F\{(i+\theta)\}] \equiv F^{(1)}(0)$, so that θ cannot be identified by $\{(X_{1i}, X_{2i})\}$.

4. SIMULATION

To compare the two different methods we undertook a small simulation study. The data were generated as follows. We let ϵ_{ji} be independent normal random variables with mean zero and standard deviation σ. For an independent uniform random variable U and for $q = 0, 1$, for $x = ih$ we defined

$$I_i(0) = \sin(3x+3h) - \sin(3x) + \sin\{(3+U)(x+h)\} - \sin\{(3+U)x\} + q\left\{(x+h)^2 - x^2\right\}\,. \qquad (4.1)$$

The choice $q = 0$ in (4.1) is meant to replicate a stationary signal, while $q = 1$ allows a nonstationary linear trend in f.

The penalized likelihood method was implemented for $r = 4$ assuming independent errors with unknown variances estimated from the data as in section 2. The maximal covariance estimators were implemented for $r = 3, 4$. We fixed $\sigma = .10$, $n = 512$ and we repeated the simulations 100 times for each of the values $h = .05, .10, .20$, $\theta = .05, .20$ and $q = 0, 1$. The results are given in Table 2, where we have evaluated median bias, RMSE (square root of the mean squared error) and MAE (median absolute error).

An inspection of Table 1 indicates that for small values of h $(= 0.05)$ none of the estimators is particularly good. This is an expected consequence of the theory, where we have assumed that $nh^4 \to \infty$: small h violates the spirit of this assumption.

For moderate and larger values of h $(= .10$ and $.20)$, with a stationary signal $(q = 0)$, all the estimators behave similarly. Because the penalized least squares estimator assumes

m–dependence with the correct choice realized ($m = 0$), we take these results as positive for the maximal covariance method. Indeed, with a more reasonable error covariance structure, e.g. $m = 5$, we would expect the penalized least squares method's performance to deteriorate. To check this, we did repeat the case $h = .20$, $q = 0$ and $\theta = .20$ with the penalized least squares approach assuming that $m = 5$. The median bias, RMSE and MAE were $-.003$, .0146, and .0092, respectively. Surprisingly, these results are as good as those with $m = 0$ known.

For the nonstationary signals ($q = 1$) the penalized least squares method emerged as clearly superior to the maximal covariance method. In simulations not presented here the maximum covariance estimator did indeed perform better if X_{1i} and X_{2i} were detrended by subtracting linear least squares fits from each process. It is interesting to note that the choice $r = 3$ in the maximum covariance method is far inferior to $r = 4$ because of bias when $r = 3$. Here we see what our theory predicts: bias reduction for larger values of r.

5. A TWO DIMENSIONAL IMAGE

As an illustration we applied these techniques to two 512×512 images provided to us by Dr. Mark Berman of CSIRO; one of the images is given as Figure 1. The pictures are aerial views of a suburb in Adelaide, Australia, recorded by a MEIS–II aircraft scanner. The other image is identical to the first by eye, although misregistration was thought to be likely, with values of (θ_1, θ_2) presumed to be in the interval $[-1, 1]^2$. In order to speed up processing time we computed (2.17) and (3.3) on a 31×31 subgrid of the image, the subgrid consisting of every 16th point in each direction.

The penalized least squares method was computed assuming 2–dependence for $r = 2, \ldots, 5$. The estimates from this method were $(-0.539, 0.134)$, $(-.472, .111)$, $(-.440, .100)$ and $(-.454, .107)$ for $r = 2, 3, 4, 5$, respectively. The maximum covariance estimator requires $r \geq 3$; for $r = 3$ it was outside the plausible interval $[-1, 1]^2$. For $r = 4, 5$, the estimates were $(-0.351, 0.092)$ and $(-0.387, 0.092)$, respectively. The functions (2.17) and (3.3) had no local optima; for example in Figure 2 we plot the negative of the cross–covariance function (3.3) when $r = 5$.

The fact that the cross–covariance estimator was unstable for $r = 3$ is in some agreement with our theory and the simulations of Section 4, both of which suggest that the cross–covariance estimator requires one extra degree of smoothness to attain the same behavior of the penalized least squares estimate.

6. DISCUSSION

We have discussed methods for estimating the lag, or amount of misregistration, in two noisy signals where the recorded data are obtained by pixel averaging. Both methods employ mathematical interpolation to accommodate the effect of the pixel averaging, and both methods use the approximation (2.9) as their starting point.

The least squares method minimizes the averaged squared distance of the signals to their interpolated estimates, but a penalty is required to take into account the errors–in–variables phenomenon, thus leading to a penalized least squares criterion. Computation of the penalty requires estimation of the covariance functions of the error processes $\{\epsilon_{1i}\}$ and $\{\epsilon_{2i}\}$, which we have assumed to be second order stationary. We have developed such estimators in the case of m-dependence, exhibiting explicit formulae in the two dimensional case when the covariance function is isotropic. If the pixel length is h and the averaged signal has r derivatives then the penalized least squares estimator of misregistration has bias of order h^{r-1} and error about the mean of order $(nh^4)^{-1/2}$.

A second method maximizes the cross–covariance function of the observed signal and its interpolated estimate. This estimator is conceptually and computationally simpler than penalized least squares, is consistent for a wide class of error processes including but not limited to m–dependence, and avoids the need to estimate the covariance function of the error processes. Its error about the mean is of order $(nh^4)^{-1/2}$, but its bias is of order $h^{\min(3,r-2)}$ and hence its bias converges to zero more slowly than does the bias of penalized least squares, for the same value of r.

For the maximal cross–covariance estimator of Section 3, rough calculations indicate that if one uses a certain value r in defining the estimators, while the real number of derivatives is $r' < r$, then one is not hurt in terms of rates of convergence, i.e., one does as well as if one assumed only r' derivatives. This issue of robustness of formulation will be pursued in a later paper.

Finally we comment of the issue of differentiability. Suppose there are a few locations where the signal is not differentiable. Then the corresponding pixel averages could, in principle, spoil the estimators, even if they represent only a small proportion of the data. This reasoning suggests that one might wish to consider robust criteria as possible replacements for our penalized least squares and maximal cross–covariance methods. For example, in the latter one might construct a robust estimator of the covariance function. We hope to pursue this issue in future work.

We believe that ours is the first formal study of rates of converge in misregistration problems with pixel averaging, and that the use of interpolation and the resulting estimators are of interest in their own right. However, many open problems remain. Among these is the development of estimators of the error process covariance functions for situations other than m–dependence. In addition the asymptotic distribution of the estimates of θ remains to be developed; the expansions (2.20) and (3.4) should prove useful in this regard.

ACKNOWLEDGEMENTS

We are most grateful to Dr. Mark Berman, of CSIRO's Division of Mathematics and Statistics, for many helpful and motivating conversations. The comments of two referees, an Associate Editor and the Editor have also been helpful, particularly in indicating places where greater clarity and stronger motivation was required. Carroll's research was supported by grants from the Institute of General Medical Sciences, the National Cancer Institute and a Visiting Fellowship to the Centre for Mathematics and its Applications at the Australian National University. Ruppert's research was supported by a grant from the National Science Foundation and a Visiting Fellowship to the Centre for Mathematics and its Applications at the Australian National University.

REFERENCES

Abramowitz, M. & Stegun, I.A. (1965). *Handbook of Mathematical Functions*. Dover, New York.

Amit, Y., Grenander, U. and Piccioni, M. (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association*, **86**, 376–387.

Anuta, P.E., Bartollucci, L.A., Dean, E., Lozano, D.F., Malaret, E., McGillem, C.D., Valdes, J.A. & Valenzuela, C.R. (1984). LANDSAT–4 MSS and thematic mapper data quality and information content analysis. *IEEE Trans. Geoscience and Remote Sensing*, **22**, 222–236.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision,* *Graphics and Image Processing* **46**, 1–21.

Berman, M., Green, A.A., Bischof, L., Davies, S.J. & Craig, M. (1990). A comparison of methods for estimating band–to–band misregistrations. In: *Proc. Fifth Australasian Remote Sensing Conference* (Perth, 8–12 October 1990), pp.987–996.

Bookstein, F.L. (1989). Principal warps: thin plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 567–585.

Fuller, W. A. (1987). *Measurement Error Models*, Wiley, New York.

Hamon, B.V. & Hannan, E.J. (1974). Spectral estimation of time delay for dispersive and non–dispersive systems. *Applied Statistics*, **23**, 134–142.

Hannan, E.J. & Thomson, P.J. (1988). Time delay estimation. *Journal of Time Series Analysis* **9**, 21–33.

Pham, D. T., Möcks, J., Köhler, W. & Gasser, T. (1987). Variable latencies of noisy signals: estimation and testing in brain potential data. *Biometrika*, 74, 525–533.

Steffensen, J.S. (1950). *Interpolation.* Chelsea, New York.

Tian, Q. & Huhns, M.N. (1986). Algorithms for subpixel registration. *Computer Vision, Graphics and Image Processing* **35**, 220-233.

Wrigley R.C., Hlavka, C.A., Card, D.H. & Buid, J.S. (1985). Evaluation of thematic mapper interband registration and noise characteristics. *Photogrammetric Engineering and Remote Sensing* **51**, 1417–1425.

CAPTION FOR TABLE 1: Constants C_{jp} and C'_{jp} from formulae (2.8).

CAPTION FOR TABLE 2: Bias, root mean squared error and median absolute error for various estimators. The study is as described in the text, and the model is given by (4.1). The value $q = 0$ refers to the stationary case, while $q = 1$ allows a nonstationary linear trend. In the table, "Bias" refers to median bias, "RMSE" is the square root of the mean squared error, and "MAE" is the median absolute error. The entry "*" indicates an abnormally large value due to a single outlier.

CAPTION FOR FIGURE 1: Aerial view of Adelaide, South Australia, recorded by a MEIS–II aircraft scanner.

CAPTION FOR FIGURE 2: Negative cross-covariance function, on vertical axis, plotted against position in square $[-1, 1] \times [-1, 1]$.

Table 1.a: $(-1)^{j+p}\,\frac{C_{jp}}{j!}$

p \ j	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	2	3	1							
4	6	11	6	1						
5	24	50	35	10	1					
6	120	274	225	85	15	1				
7	720	1764	1624	735	175	21	1			
8	5040	13068	13132	679	1960	322	28	1		
9	40320	109584	118124	67284	22449	4536	546	36	1	
10	362880	1026576	1173680	723680	269325	63273	9450	870	45	1

Table 1.b: $\frac{C'_{jp}}{j!}$

p \ j	1	2	3	4	5	6	7	8	9	10
1	1									
2	0	1								
3	$-\frac{1}{4}$	0	1							
4	0	-1	0	1						
5	$\frac{9}{16}$	0	$\frac{5}{2}$	0	1					
6	0	4	0	-5	0	1				
7	$-\frac{225}{64}$	0	$\frac{259}{16}$	0	$-\frac{35}{4}$	0	1			
8	0	-36	0	49	0	-14	0	1		
9	$\frac{11025}{256}$	0	$-\frac{3229}{8}$	0	$-\frac{987}{8}$	0	-21	0	1	
10	0	576	0	-820	0	293	0	-30	0	1

Table 2: Results of a Simulation Study

h	q	θ	Penalized Least Squares Bias	Penalized Least Squares RMSE	Penalized Least Squares MAE	Maximal Covariance, r=3 Bias	Maximal Covariance, r=3 RMSE	Maximal Covariance, r=3 MAE	Maximal Covariance, r=4 Bias	Maximal Covariance, r=4 RMSE	Maximal Covariance, r=4 MAE
.05	0	.05	−.039	.520	.226	−.047	2.04	.268	.002	1.31	.280
.05	1	.05	.082	.700	.243	.456	3.98	.675	−.027	1.62	.287
.10	0	.05	.0023	.072	.041	.019	.064	.037	.001	.068	.036
.10	1	.05	.007	.084	.056	.159	.289	.171	.011	.107	.062
.20	0	.05	.001	.0148	.0096	.001	.0146	.0097	.001	.0146	.0094
.20	1	.05	.0007	.0127	.0079	.159	.203	.159	.003	.0230	.014
.05	0	.20	.129	.552	.270	.124	2.92	.353	.126	26.9	.287
.05	1	.20	−.201	.769	.276	.258	8.41	.722	−.151	4.70	.443
.10	0	.20	.004	.071	.046	.002	.064	.042	.007	.067	.046
.10	1	.20	.020	*	.059	.160	*	.198	.004	*	.064
.20	0	.20	.003	.0162	.0120	.000	.0155	.0111	.003	.0156	.0116
.20	1	.20	.000	.0146	.0100	.148	.186	.148	.007	.0264	.0181