CIRCULAR REGRESSION USING BAYESIAN UNWRAPPING

Mark R. Morelande

Melbourne Systems Laboratory
Department of Electrical and Electronic Engineering
The University of Melbourne, Australia
email: m.morelande@ee.unimelb.edu.au

ABSTRACT

Circular data arise in a number of biological and physical applications. Circular regression refers to the study of the dependence of a circular response variable on a collection of explanatory variables. In this paper the circular response variable is modelled as a wrapped Gaussian process. Previously, estimation with wrapped processes has been performed used complicated iterative optimisation or random sampling techniques. The recursive Bayesian algorithm proposed here is simple to implement and computationally economical by comparison. The proposed algorithm is applied to phase parameter estimation.

Keywords: Parameter estimation, Bayes procedures.

1. INTRODUCTION

Linear regression is a well-known procedure used to investigate the dependence of a response variable on a number of explanatory variables. Mathematically, assume n measurements of the response variable $y_1, \ldots, y_n$, $y_t \in \mathbb{R}$ and corresponding explanatory variables $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_n$, $\boldsymbol{h}_t \in \mathbb{R}^{1\times r}$ which satisfy

$$y_t = \boldsymbol{h}_t\boldsymbol{\theta} + e_t, \quad t = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\theta} \in \mathbb{R}^r$ is an unknown vector parameter and $e_1, \ldots, e_n$ are zero-mean random variables. The first element of $\boldsymbol{h}_t$ is taken to be one. The problem of linear regression is to estimate $\boldsymbol{\theta}$.

There are certain cases of interest in which the response variable is periodic. This occurs, for instance, if one measures directions or the times of events during a cycle [7]. Such observations are said to be circular since they can be represented as a point on the circumference of a circle. Analogously to linear regression, the goal of circular regression is to investigate the dependence of a circular response variable on explanatory variables. Circular regression is used in a number of physical and biological applications including studies of the dependence of wind direction on ozone level [8] and of the dependence of direction on distance travelled in animal migration [4]. More examples can be found in [7].

The von Mises distribution is a natural way to model circular variables and has been widely used in circular regression [4, 6, 8]. Circular variables can also be modelled as wrapped processes [5, 10]. This will be the approach taken here. The circular response variables $x_1, \ldots, x_n$ are then modelled as

$$x_t = y_t \bmod 2\pi, \quad t = 1, \ldots, n, \tag{2}$$

where y_t is given in (1). The problem is to estimate the parameter $\boldsymbol{\theta}$ given $\boldsymbol{x} = [x_1, \ldots, x_n]'$ where $'$ denotes matrix transpose. It will be assumed that $e_1, \ldots, e_n$ are independent Gaussian random variables with variance ρ^2.

Estimation with the wrapped processes has previously been considered in [5, 10]. Both algorithms use the equivalent signal model $y_t = x_t + 2\pi k_t$ where $k_1, \ldots, k_n$ are unknown integers, referred to as unwrapping variables. In [5] the expectation-maximisation (EM) algorithm is used with $\boldsymbol{x}$ the incomplete, observed data and the unwrapping variables the missing data. Application of EM is complicated due to the need to evaluate infinite summations. In [10] the idea of data augmentation, with x_t, k_t, $t = 1, \ldots, n$ the augmented data, was used to develop a Markov chain Monte Carlo technique for estimating $\boldsymbol{\theta}$.

The approach taken here is somewhat simpler than those of [5, 10]. The key idea is that, conditional on a sequence of unwrapping variables, $\boldsymbol{\theta}$ can be optimally estimated in closed-form. In order to apply this idea it is necessary to determine which of the infinite number of possible unwrapping sequences are most probably correct. This is done by adopting a Bayesian framework in which the validity of a particular unwrapping sequence can be measured by its posterior probability. Since it is not possible to compute the posterior probabilities of an infinite number of unwrapping sequences a recursive procedure is developed to eliminate unlikely sequences. The proposed algorithm is simple to implement and computationally undemanding. Although the Bayesian framework permits it, there is no need to have accurate prior information regarding $\boldsymbol{\theta}$. The performance of the algorithm is demonstrated for the problem of phase parameter estimation.

The paper is organised as follows. The exact posterior distribution is derived in Section 2 and a scheme for approximating it is developed in Section 3. Several issues related to the proposed algorithm are discussed in Section 4 and an example application is given in Section 5.

2. EXACT BAYESIAN ESTIMATION

The goal of Bayesian estimation is to compute the posterior probability density function (PDF) of the parameter of interest, in this case $\boldsymbol{\theta}$. Once the posterior PDF is available quantities of interest related to $\boldsymbol{\theta}$ can, in principle, be computed. In particular, the posterior mean is an optimal estimator of $\boldsymbol{\theta}$ in the mean square error sense. The exact posterior PDF, and hence the posterior mean, for the circular regression problem will be derived in this section. The notation $N(\boldsymbol{z}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ will be used to denote the Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{z}$.

The basic elements of Bayesian estimation are a prior PDF for $\boldsymbol{\theta}$ and the likelihood of $\boldsymbol{\theta}$ given the observations $\boldsymbol{x}$. The prior PDF will be taken to be

$$p(\boldsymbol{\theta}) = N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_0, \boldsymbol{\Sigma}_0), \tag{3}$$

where $\hat{\boldsymbol{\theta}}_0$ and $\boldsymbol{\Sigma}_0$ are the prior mean and covariance matrix, re-

spectively. The likelihood can be found, using (2) and (1), as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^n} N(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{\theta} - 2\pi\boldsymbol{k}, \rho^2\boldsymbol{I}_n), \tag{4}$$

where $\mathbb{Z} = \{0, \pm1, \pm2, \ldots\}$ and $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. Using Bayes' rule, (3) and (4), the posterior PDF can be written as

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{x}) &\propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{k}\in\mathbb{Z}^n} N(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{\theta} - 2\pi\boldsymbol{k}, \rho^2\boldsymbol{I}_n) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_0, \boldsymbol{\Sigma}_0) \end{aligned} \tag{5}$$

The following lemma is used to evaluate (5).

Lemma 1 Assume that $\boldsymbol{z} \in \mathbb{R}^d$, $\boldsymbol{A} \in \mathbb{R}^{d\times m}$, $\boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^m$ and that $\boldsymbol{B} \in \mathbb{R}^{d\times d}$ $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$ are positive definite matrices. Then,

$$N(\boldsymbol{z}; \boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}) N(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = N(\boldsymbol{z}; \boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{S}) N(\boldsymbol{x}; \boldsymbol{\nu}, \boldsymbol{\Omega}), \tag{6}$$

where $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}' + \boldsymbol{B}$ and, with $\boldsymbol{K} = \boldsymbol{\Sigma}\boldsymbol{A}'\boldsymbol{S}^{-1}$,

$$\begin{aligned} \boldsymbol{\nu} &= \boldsymbol{\mu} + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{A}\boldsymbol{\mu}), \\ \boldsymbol{\Omega} &= \boldsymbol{\Sigma} - \boldsymbol{K}\boldsymbol{A}\boldsymbol{\Sigma}. \end{aligned}$$

Proof: See [2].

Let $\hat{\boldsymbol{x}}(\boldsymbol{k}) = \boldsymbol{H}\hat{\boldsymbol{\theta}}_0 - 2\pi\boldsymbol{k}$, $\boldsymbol{k} \in \mathbb{Z}^n$, $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{\Sigma}_0\boldsymbol{H}' + \rho^2\boldsymbol{I}_n$ and $\boldsymbol{G} = \boldsymbol{\Sigma}_0\boldsymbol{H}'\boldsymbol{S}^{-1}$. Then, applying Lemma 1 to (5) allows the posterior PDF to be written as

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^n} \gamma_n(\boldsymbol{k}) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_n(\boldsymbol{k}), \boldsymbol{\Sigma}_n) \tag{7}$$

where, for $\boldsymbol{k} \in \mathbb{Z}^n$,

$$\gamma_n(\boldsymbol{k}) = N(\boldsymbol{x}; \hat{\boldsymbol{x}}(\boldsymbol{k}), \boldsymbol{S}) \Big/ \sum_{\boldsymbol{d}\in\mathbb{Z}^n} N(\boldsymbol{x}; \hat{\boldsymbol{x}}(\boldsymbol{d}), \boldsymbol{S}) \ , \tag{8}$$

$$\hat{\boldsymbol{\theta}}_n(\boldsymbol{k}) = \hat{\boldsymbol{\theta}}_0 + \boldsymbol{G}[\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{k})], \tag{9}$$

$$\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0 - \boldsymbol{G}\boldsymbol{H}\boldsymbol{\Sigma}_0. \tag{10}$$

The quantity $\gamma_n(\boldsymbol{k})$ in (7) is the posterior probability of the unwrapping sequence $\boldsymbol{k} \in \mathbb{Z}^n$. Conditional on this unwrapping sequence, the posterior distribution of $\boldsymbol{\theta}$ is Gaussian with mean $\hat{\boldsymbol{\theta}}_n(\boldsymbol{k})$ and covariance matrix $\boldsymbol{\Sigma}_n$. The posterior mean of $\boldsymbol{\theta}$ can be obtained from (7) as

$$\hat{\boldsymbol{\theta}} = \sum_{\boldsymbol{k}\in\mathbb{Z}^n} \gamma_n(\boldsymbol{k})\hat{\boldsymbol{\theta}}_n(\boldsymbol{k}). \tag{11}$$

Eq. (11) can be thought of as a probabilistic unwrapping of the circular observations, with estimates from each possible unwrapping sequence weighted according to the posterior probability of the unwrapping sequence. An important aspect of this estimator is that the conditional means $\hat{\boldsymbol{\theta}}_n(\boldsymbol{k})$ can be computed in closed-form.

An obvious problem with the estimator (11) is that it is not possible to enumerate the countably infinite number of unwrapping sequences. However a practical estimator can be realised by considering only a relatively small number of unwrapping sequences. A recursive procedure for constructing the viable set of unwrapping sequences will be described in the following section. The motivation behind using a recursive procedure is that it allows unlikely unwrapping sequences to be discarded based on the observations which have been processed.

3. APPROXIMATE BAYESIAN ESTIMATION

Throughout this section the notation $\boldsymbol{x}_t = [x_1, \ldots, x_t]'$ will be used to denote the vector of the first t observations. Similar notation will be used for unwrapping sequences. Using Bayes' rule the posterior PDF of $\boldsymbol{\theta}$ using only the first measurement is

$$p(\boldsymbol{\theta}|x_1) \propto \sum_{k_1=-\infty}^{\infty} N(x_1; \boldsymbol{h}_1\boldsymbol{\theta} - 2\pi k_1, \rho^2) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_0, \boldsymbol{\Sigma}). \tag{12}$$

Eq. (12) has an infinite number of terms and so cannot be evaluated. However many of the terms are negligible and can be ignored for practical purposes. The viability of an unwrapping variable can be assessed by considering the prior measurement PDF,

$$p(x_1, k_1) = \int p(x_1, k_1|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} = N(x_1; \hat{x}_1(k_1), s_1), \tag{13}$$

where $\hat{x}_1(k_1) = \boldsymbol{h}_1\hat{\boldsymbol{\theta}}_0 - 2\pi k_1$ and $s_1 = \boldsymbol{h}_1\boldsymbol{\Sigma}_0\boldsymbol{h}_1' + \rho^2$. Eq. (13) suggests selecting the set of viable unwrapping variables for $t = 1$ as

$$\tilde{K}_1 = \{k_1 : |x_1 - \hat{x}_1(k_1)| < \sqrt{s_1}\Phi^{-1}(1 - \alpha/2)\} \tag{14}$$

where Φ is the distribution function of a standard Gaussian random variable and α is the probability of excluding the correct unwrapping sequence. The posterior PDF is then approximated by

$$\tilde{p}(\boldsymbol{\theta}|x_1) \propto \sum_{k_1\in\tilde{K}_1} N(x_1; \boldsymbol{h}_1\boldsymbol{\theta} - 2\pi k_1, \rho^2) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_0, \boldsymbol{\Sigma}) \tag{15}$$

Using Lemma 1 gives

$$\tilde{p}(\boldsymbol{\theta}|x_1) = \sum_{k_1\in\tilde{K}_1} \tilde{\gamma}_1(k_1) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_1(k_1), \boldsymbol{\Sigma}_1) \tag{16}$$

where, with $\boldsymbol{g}_1 = \boldsymbol{\Sigma}_0\boldsymbol{h}_1'/s_1$ and $\psi_1(k_1) = N(x_1; \hat{x}_1(k_1), s_1)$,

$$\tilde{\gamma}_1(k_1) = \psi_1(k_1) \Big/ \sum_{d\in\tilde{K}_1} \psi_1(d) \ , \tag{17}$$

$$\hat{\boldsymbol{\theta}}_1(k_1) = \hat{\boldsymbol{\theta}}_0 + \boldsymbol{g}_1[x_1 - \hat{x}_1(k_1)], \tag{18}$$

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_0 - \boldsymbol{g}_1\boldsymbol{h}_1\boldsymbol{\Sigma}_0. \tag{19}$$

The use of (14) to remove from consideration unlikely unwrapping variables avoids the need to evaluate an infinite number of unwrapping sequences. However, if sets of valid unwrapping sequences are constructed simlarly to (14) for successive response variables, the number of sequences considered after t steps will be exponential in t. Since this is computationally infeasible for any reasonable t further reduction of the mixture (16) is required. The following simple pruning scheme is used for this purpose. Order the posterior probability approximations (17) as $\tilde{\gamma}_1^{(1)} > \tilde{\gamma}_1^{(2)} > \cdots > \tilde{\gamma}_1^{(\tilde{m})}$ where $\tilde{m} = |\tilde{K}_1|$. The set $\tilde{K}_1$ of viable unwrapping variables is reduced to

$$K_1 = \{k_1 \in \tilde{K}_1 : \tilde{\gamma}_1(k_1) \geq \epsilon\} \tag{20}$$

where $\epsilon = \max(\tilde{\gamma}_1^{(m)}, \epsilon_0)$ if $\tilde{m} > m$ and $\epsilon = \epsilon_0$ if $\tilde{m} \leq m$. The parameter m is the maximum allowable number of unwrapping sequences and $\epsilon_0 \in [0, 1)$ is a user-selected threshold. Then, the final approximation to the posterior PDF for $t = 1$ is

$$\hat{p}(\boldsymbol{\theta}|x_1) = \sum_{k_1\in K_1} \hat{\gamma}_1(k_1) N(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_1(k_1), \boldsymbol{\Sigma}_1) \tag{21}$$

where $\hat{\gamma}_1(k_1) = \tilde{\gamma}_1(k_1)/\sum_{d\in K_1}\tilde{\gamma}_1(d)$. The procedure of (14) and (17)-(20) is performed sequentially for $t = 2,\ldots,n$ to obtain approximations of the posterior PDFs $p(\boldsymbol{\theta}|\boldsymbol{x}_t)$. The main difference for $t = 2,\ldots,n$ compared to $t = 1$ is that the PDF $p(\boldsymbol{\theta}|\boldsymbol{x}_{t-1})$, which is the prior PDF at the tth recursion, will be a Gaussian mixture rather than a single Gaussian. Thus it is necessary to contruct sets as in (14) for each component of the mixture with the preliminary set of viable unwrapping sequences the union of these sets. The complete procedure, to be referred to as the Bayesian unwrapping algorithm (BUA), is summarised in Table 1.

Table 1: Recursive procedure for approximating the posterior mean.

1. Set $K_0 = 0$, $\hat{\gamma}_0(0) = 1$, $\hat{\boldsymbol{\theta}}_0(0) = \hat{\boldsymbol{\theta}}_0$.
2. For $t = 1,\ldots,n$:
   (a) Compute $s_t = \boldsymbol{h}_t\boldsymbol{\Sigma}_{t-1}\boldsymbol{h}_t' + \rho^2$, $\boldsymbol{g}_t = \boldsymbol{\Sigma}_{t-1}\boldsymbol{h}_t'/s_t$.
   (b) For $\boldsymbol{k}_{t-1} \in K_{t-1}$, construct the set
   $$\tilde{K}_t(\boldsymbol{k}_{t-1}) = \{\boldsymbol{k}_t : |x_t - \hat{x}_t(\boldsymbol{k}_t)| < \sqrt{s_t}\Phi^{-1}(1-\alpha/2)\}$$
   where $\hat{x}_t(\boldsymbol{k}_t) = \boldsymbol{h}_t\hat{\boldsymbol{\theta}}_{t-1}(\boldsymbol{k}_{t-1}) - 2\pi k_t$.
   (c) Set
   $$\tilde{K}_t = \bigcup_{\boldsymbol{k}_{t-1}\in K_{t-1}} \tilde{K}_t(\boldsymbol{k}_{t-1})$$
   (d) For $\boldsymbol{k}_t \in \tilde{K}_t$, compute
   $$\psi_t(\boldsymbol{k}_t) = N(x_t; \hat{x}_t(\boldsymbol{k}_t), s_t)\hat{\gamma}_{t-1}(\boldsymbol{k}_{t-1}),$$
   $$\hat{\boldsymbol{\theta}}_t(\boldsymbol{k}_t) = \hat{\boldsymbol{\theta}}_{t-1}(\boldsymbol{k}_{t-1}) + \boldsymbol{g}_t[x_t - \hat{x}_t(\boldsymbol{k}_t)].$$
   (e) For $\boldsymbol{k}_t \in \tilde{K}_t$, normalise the posterior probabilities
   $$\tilde{\gamma}_t(\boldsymbol{k}_t) = \psi_t(\boldsymbol{k}_t) \Big/ \sum_{\boldsymbol{d}_t\in\tilde{K}_t} \psi_t(\boldsymbol{d}_t)\ .$$
   (f) Construct the set $K_t = \{\boldsymbol{k}_t \in \tilde{K}_t : \tilde{\gamma}_t(\boldsymbol{k}_t) \geq \epsilon\}$.
   (g) For $\boldsymbol{k} \in K_t$, re-normalise the posterior probabilities
   $$\hat{\gamma}_t(\boldsymbol{k}_t) = \tilde{\gamma}_t(\boldsymbol{k}_t) \Big/ \sum_{\boldsymbol{d}_t\in K_t} \tilde{\gamma}_t(\boldsymbol{d}_t)\ .$$
   (h) Compute $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_{t-1} - \boldsymbol{g}_t\boldsymbol{h}_t\boldsymbol{\Sigma}_{t-1}$.
3. Compute the estimate
   $$\hat{\boldsymbol{\theta}} = \sum_{\boldsymbol{k}_n\in K_n} \hat{\gamma}_n(\boldsymbol{k}_n)\hat{\boldsymbol{\theta}}_n(\boldsymbol{k}_n).$$

4. DISCUSSION

In this section several points of interest regarding the BUA proposed in Section 3 will be discussed. These are enumerated below.

1. In order to ensure identifiability of the parameter $\boldsymbol{\theta}$ it is necessary to restrict its range. For example, the constant offset must satisfy $\alpha \leq \theta_1 < 2\pi + \alpha$ for some $\alpha \in \mathbb{R}$. More generally the range of allowable values of $\boldsymbol{\theta}$ will depend on the explanatory variables $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_n$.

2. The BUA requires specification of a prior PDF for $\boldsymbol{\theta}$ (3). This does not prevent the BUA being used for estimation of deterministic parameters provided that the effective support of the prior covers the range of allowable parameter values. The prior has little effect on algorithm performance if the number n of observations is reasonably large.

3. Nonlinear estimators are subject to thresholding whereby the estimator variance increases dramatically for noise levels above a threshold level. Since linear estimators, such as $\hat{\boldsymbol{\theta}}_n(\boldsymbol{k}_n)$, do not exhibit thresholding, it is evident from (11) that thresholding will occur solely due to an inability to unambiguously unwrap the circular response variables via the posterior probabilities $\gamma_n(\boldsymbol{k}_n)$.

4. The posterior mean, or its approximation via Table 1, can be unduly influenced by lowly, but still appreciably weighted unwrapping sequences. This is particularly so for large noise variances and/or small sample lengths. In such cases it has been found more reliable to use the estimator $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_n(\boldsymbol{k}_n^*)$ where

$$\boldsymbol{k}_n^* = \arg\max_{\boldsymbol{k}_n\in K_n} \hat{\gamma}_n(\boldsymbol{k}_n). \tag{22}$$

5. EXAMPLE: PHASE PARAMETER ESTIMATION

The proposed algorithm will be applied to an important problem in signal processing, that of estimating the parameters of a phase modulated signal embedded in additive noise. The complex-valued observations are modelled by, for $t = 1,\ldots,n$,

$$z_t = \exp(j\boldsymbol{h}_t\boldsymbol{\theta}) + w_t, \tag{23}$$

where $w_1,\ldots,w_n$ are independent complex-valued Gaussian random variables with $\mathsf{E}(w_t) = 0$, $\mathsf{E}(w_t^2) = 0$ and $\mathsf{E}|w_t|^2 = \sigma^2$. The row vectors $\boldsymbol{h}_1,\ldots,\boldsymbol{h}_n$ contain the values of the basis functions used to fit the phase. Polynomials are a commonly used basis in which case $\boldsymbol{h}_t = [1, t, \ldots, t^q]$ for a qth order polynomial and (23) is referred to as a polynomial phase signal (PPS). The problem is to estimate $\boldsymbol{\theta}$ from the observations $z_1,\ldots,z_n$.

In [11] it is shown that, for sufficiently high SNR,

$$\angle z_t \approx y_t \bmod 2\pi, \quad t = 1,\ldots,n, \tag{24}$$

where y_t satisfies (1) with $\rho^2 = \sigma^2/2$. According to (24) the BUA can be applied to the time series $\angle z_1,\ldots,\angle z_n$ to estimate the parameter $\boldsymbol{\theta}$. A better approximation can be obtained by using the true variance,

$$\rho^2 = \mathsf{var}\left[\arctan\left(\frac{V}{1+U}\right)\right] \tag{25}$$

where $V, U \sim N(0, \sigma^2/2)$ are independent. To ensure identifiability we must have $\boldsymbol{\theta} \in \Theta = \prod_{i=0}^{q}(-\pi/i!, \pi/i!]$. The prior PDF is chosen so that parameters in Θ are reasonably well supported. The prior mean is $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{0}$ and the prior covariance matrix is $\boldsymbol{\Sigma}_0 = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_{q+1}^2)$ where $\sigma_{i+1}^2 = (\pi/(4i!))^2$, $i = 0,\ldots,q$. As discussed in Section 4, the posterior mean conditional on the most probable unwrapping sequence will be used rather than a weighted sum of posterior means as in Table 1.

The proposed algorithm will be compared with existing methods. Many algorithms have been proposed in the literature for parameter estimation of PPS. Maximum likelihood (ML) estimates can be computed via a q-dimensional iterative optimisation. The main concern is that a reasonably accurate initialisation is required

to avoid converging to a local maximum. The method of [1] alleviates but does not remove this requirement. Prominent alternatives to ML estimation include the polynomial phase transform (PPT) [9] and the phase unwrapping least-squares algorithm (PULSA) of [3]. Both of these methods use the result that differencing reduces the order of a polynomial by one. In the PPT this results in a scheme in which the q-dimensional optimisation of the ML estimate is replaced by a series of one-dimensional optimisations to successively compute estimates of $\theta_{q+1}, \theta_q, \ldots, \theta_1$. The PULSA uses the approximation (24) and the fact that the qth differenced phase is equal to a constant plus noise to unwrap the phase.

The performances of the BUA, the PPT and the PULSA will be compared for a cubic phase with $n = 100$ observations. Two experiments are performed. In the first experiment,

$$\boldsymbol{\theta} = \begin{bmatrix} \pi/5 & \pi/4 & -2\pi \times 10^{-3} & -5\pi \times 10^{-5} \end{bmatrix}'. \quad (26)$$

The parameters in (26) are selected to be within the dynamic range of the PPT, which is much smaller than those of the BUA and PULSA. By necessity some of these parameters are close to zero which is the prior mean for the BUA. Although this may appear favourable to the BUA it is not really because the prior covariance matrix is quite large meaning that little prior confidence is placed in the prior mean. To demonstrate this a second experiment in which all parameters are far from zero is performed:

$$\boldsymbol{\theta} = \begin{bmatrix} \pi/5 & \pi/4 & -\pi/10 & -\pi/18 \end{bmatrix}'. \quad (27)$$

Note that the PPT cannot be used to estimate these parameters. The signal-to-noise ratio (SNR), defined as $-10\log_{10}(\sigma^2)$, is varied between 0 dB and 20 dB. Estimator standard deviations are computed over 1000 realisations for each SNR. The results for both experiments are shown in Figure 1 for the parameter θ_4. The Cramér-Rao bound (CRB) is also shown. The two important aspects of algorithm performance are the threshold SNR and the accuracy at SNRs above the threshold SNR. For both experiments the BUA has by far the lowest threshold SNR. In experiment one, the BUA threshold is 6 dB lower than the PPT threshold and 12 dB lower than the PULSA threshold. In experiment two it is 15 dB lower than the PULSA threshold. The BUA and the PULSA perform equally above their respective thresholds with the standard deviations of both estimators achieving the CRB at high SNRs. The standard deviation of the BUA is slightly larger than the CRB for small SNRs due to the use of the high SNR approximation (24).

The computational expense of the BUA depends on the SNR. Although the BUA is more computationally demanding than the PPT and PULSA for all SNRs, the difference is small for high SNRs, when the phase can be unwrapped with little ambiguity, and large for low SNRs, when phase unwrapping can be done reliably only by retaining many candidate unwrapping sequences. A major part of the difficulty in unwrapping the phase at low SNRs is a break down of the Gaussian assumption. In particular, for low SNRs, the noise in $\angle z_t$ follows a distribution with a heavier tail than the Gaussian distribution. The resulting occurence of unmodelled outliers makes phase unwrapping difficult.

6. CONCLUSIONS

The problem of circular regression, in which the aim is to quantify the dependence of a circular response variable on explanatory variables, was considered. The algorithm proposed here uses a wrapped Gaussian process to model the circular response variables. The principal concern in such a model is to unwrap the circular variables. A recursive Bayesian algorithm was proposed for this purpose. The proposed algorithm was successfully applied to the problem of phase parameter estimation. An important area for future work is the extension of the algorithm to estimation of the scaling parameter.

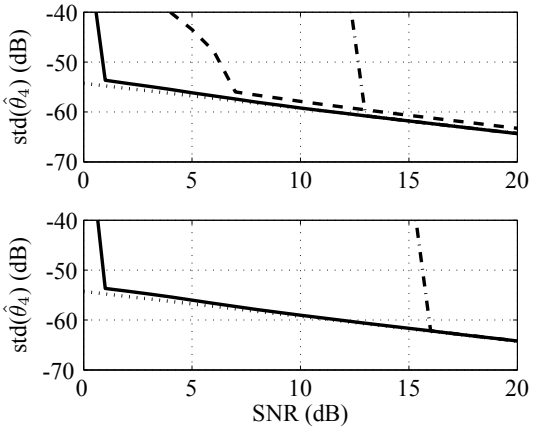

Figure 1: Standard deviations of the BUA (solid), PPT (dashed) and PULSA (dash-dot) estimators of θ_4 plotted against SNR for experiment one (top) and experiment two (bottom). The dotted line is the CRB. The signal has a cubic phase and $n = 100$.

7. REFERENCES

[1] J. Ängeby, "Estimating signal parameters using the nonlinear instantaneous least squares approach," *IEEE Transactions on Signal Processing*, vol. 48, no. 10, pp. 2721–2732, October 2000.

[2] S. Challa and D. Koks, "Bayesian and Dempster-Shafer fusion," *Sadhana*, vol. 29, no. 2, pp. 145–176, 2004.

[3] P. Djurić and S. Kay, "Parameter estimation of chirp signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 12, pp. 2118–2126, 1990.

[4] N. Fisher and A. Lee, "Regression models for an angular response," *Biometrics*, vol. 48, no. 3, pp. 665–677, 1992.

[5] ——, "Time series analysis of circular data," *Journal of the Royal Statistical Society. Series B*, vol. 56, no. 2, pp. 327–339, 1994.

[6] A. Gould, "A regression technique for angular variates," *Biometrics*, vol. 25, no. 4, pp. 683–700, 1969.

[7] S. Jammalamadaka and A. SenGupta, *Topics in Circular Statistics*. Singapore: World Scientific, 2001.

[8] R. Johnson and T. Wehrly, "Some angular-linear distributions and related regression models," *Journal of the American Statistical Association*, vol. 73, no. 3, pp. 602–606, 1978.

[9] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1901–1914, 1995.

[10] P. Ravindran, "Bayesian analysis of circular data using wrapped distributions," Ph.D. dissertation, North Carolina State University, 2002.

[11] S. Tretter, "Estimating the frequency of a noisy sinusoid by linear prediction," *IEEE Transactions on Information Theory*, vol. 31, pp. 832–835, November 1985.