PARAMETER ESTIMATION OF LINEAR FREQUENCY MODULATED SIGNALS WITH MISSING OBSERVATIONS

Ramon Brcich and Abdelhak M. Zoubir

Communications and Signal Processing Group
Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia.
Email : {r.brcich,zoubir}@ieee.org

ABSTRACT

The effect of missing observations, caused by random sensor errors or periodic blocking of the signal, on the estimation of signal parameters is to reduce estimator performance. For the class of polynomial phase signals, existing estimators may suffer a severe degradation in performance depending on the location of the missing observations. Reasons for the drop in performance are determined and several estimators are proposed which behave well even with a large proportion of missing observations.

1. INTRODUCTION

The problem of missing observations occurs when a signal becomes unavailable. This may occur randomly due to path loss or sensor failure, and in blocks due to periodically available astronomical observations or misses in passive electronic intelligence measurements due to antenna sweep-back [1].

Polynomial phase signals (PPS) are suitable models in many of these applications as they account for the frequency modulation induced by movement, in synthetic aperature radar the signal is usually assumed to be a linear frequency modulated (LFM) or chirp signal.

Existing estimators for the parameters of PPS can be quite sensitive to even a small amount of missing observations necessitating the design of estimators which perform well even with a large number of missing observations. Here, existing estimators for the parameters of LFM signals are assessed and several improved algorithms are proposed.

2. SIGNAL MODEL

Consider a nonstationary complex analytic signal $z(t)$ sampled at the times $t \in T = \{t_n, n = 1, \dots, N\}$,

$$z(t_n) = b(t_n)As(t_n;\boldsymbol{a}) + w(t_n), \qquad n = 1, \dots, N, \qquad (1)$$

where the $b(t_n) \in \{0, 1\}$ indicate the missing observations, A is a constant amplitude, $s(t_n;\boldsymbol{a})$ is the signal of interest with parameter vector $\boldsymbol{a}$ and $w(t_n)$ are independent and identically distributed (i.i.d) complex circular Gaussian random variables with zero mean and variance σ^2. The problem is to estimate $\boldsymbol{\theta} = (\sigma^2, A, \boldsymbol{a}^\mathsf{T})^\mathsf{T}$.

Here, $s(t_n;\boldsymbol{a})$ is a polynomial phase signal (PPS),

$$s(t_n;\boldsymbol{a}) = \exp\left(\jmath \sum_{m=0}^{M} a_m t_n^m\right), \qquad \boldsymbol{a} = (a_0, \dots, a_M)^\mathsf{T}, \quad (2)$$

This work was supported under the ARC (Australian Research Council) Discovery Project scheme.

where M is the order of the PPS. Only the case $M = 2$ is considered, so that $s(t_n)$ (the reference to $\boldsymbol{a}$ will be dropped when the meaning is clear) is a LFM signal. Herein the sampling is uniform in time, $t_{n+1} - t_n = \Delta$ where the sampling period $\Delta = 1$.

The approach taken to solving the problem (1) depends on the characteristics of $b(t_n)$. Two general cases will be considered, randomly missed observations, and blocks of missing observations. For randomly missed observations the $b(t_n)$ are modeled as i.i.d Bernoulli random variables, this is denoted model b_0,

Model b_0: $\Pr(b(t_n) = 1) = 1 - p, \quad \Pr(b(t_n) = 0) = p,$

where p is the probability of a missing observation. Contiguous blocks of missing observations are unlikely in the random misses model, but are likely in reality. Two possibilities are considered,

Model b_1: $[Np]$ observations are missing from the start of the signal, $[\cdot]$ denoting the integer part,

Model b_2: $[Np]$ observations are missing from the centre of the signal.

These two choices were found to represent well the behaviour of estimators for blocks of missing observations.

A distinction may be made regarding knowledge of $b(t_n)$. If $b(t_n)$ and hence the locations of the missing observations are known, the problem can be viewed as one of irregular or nonuniform sampling. The more difficult problem occurs when the $b(t_n)$ are unknown. In either case $N(1 - p)$ observations contain information about A and $\boldsymbol{a}$ on the average.

3. BEHAVIOUR OF EXISTING ESTIMATORS

Several estimators for the parameters of PPS exist including maximum likelihood estimators (MLE) [2], least squares regression based on phase unwrapping [3], delay and conjugation based operators such as the discrete polynomial phase transform (DPT) [4] and extensions to the DPT utilising the ambiguity function. The performance of the MLE and DPT are considered here for several reasons. First, the MLE is computationally feasible for LFM signals and is the optimal estimator in terms of variance. Second, the DPT is a popular method due to its low computational complexity and reasonable performance for lower polynomial phase orders. Methods based on phase unwrapping were found to be very sensitive to even a small proportion ($< 5\%$) of missing observations, while most other methods have a computational complexity and performance somewhere between the MLE and DPT.

Any estimator will suffer a degradation in performance as the number of missing observations increase since the amount of information available about the parameters is decreased. Next, the performance of the MLE and the DPT for LFM signals with missing observations is considered in more detail.

3.1. Maximum Likelihood Estimation

A computationally efficient MLE algorithm for LFM signals based on a coarse grid search and a fine Newton-Raphson search was developed in [2]. This is the MLE referred to herein.

3.1.1. Variance

A guide to minimum increase in the Cramér-Rao bound (CRB) when observations are missing may be obtained by determining the CRB for the parameters of interest given that the locations of the missing observations are known.

First consider the Fisher information matrix (FIM) $\boldsymbol{J}$ for the case of no missed observations, that is, $p = 0$. The structure of $\boldsymbol{J}$ reveals $\widehat{\sigma^2}$, $\hat{A}$ and $\hat{\boldsymbol{a}}$ to be asymptotically mutually independent,

$$\boldsymbol{J} = \left[\begin{array}{cc|c} J_{\sigma^2\sigma^2} & 0 & \\ 0 & J_{AA} & \\ \hline & & \boldsymbol{J_{aa}} \end{array}\right], \quad (3)$$

where $J_{\sigma^2\sigma^2} = N/\sigma^4$, $J_{AA} = 2N/\sigma^2$ and $(\boldsymbol{J_{aa}})_{ij} = 2A^2/\sigma^2 \sum_{n=1}^{N} t_n^{i+j-2}$.

To proceed, use will be made of the result [5] that for random sampling schemes each element of the FIM, $(\boldsymbol{J}')_{ij}$, is found by taking the expectation of each element of the FIM $(\boldsymbol{J}(T'))_{ij}$, a deterministic function of the sampling times T', over all possible sets $T' \subseteq T$ of the sampling times.

This result must be applied twice, first to find the information present in $z(t_n)$ when $b(t_n) = 1$ and second to find the information present when $b(t_n) = 0$. The reason for this is that information about σ^2 is present in all observations regardless of $b(t_n)$, while information about the remaining parameters is only present if $b(t_n) = 1$. As the noise is i.i.d, the general expression for the FIM when observations are missing is

$$\boldsymbol{J}_{b_i} = \sum_{n=1}^{N} \boldsymbol{J}(t_n|b(t_n) = 0)\mathsf{E}\left[I(b(t_n) = 0)\right] + \boldsymbol{J}(t_n|b(t_n) = 1)\mathsf{E}\left[I(b(t_n) = 1)\right] \quad (4)$$

where $I(\cdot)$ is the indicator function and $i = 1, 2, 3$ denotes the three models for the location of missing observations.

For model b_0, with Bernoulli random missed observations, each element of the first term of (4) is zero except for that corresponding to σ^2, which is $pJ_{\sigma^2\sigma^2}$. The second term evaluates to $(1 - p)\boldsymbol{J}$ giving the FIM as

$$\boldsymbol{J}_{b_0} = \left[\begin{array}{cc|c} J_{\sigma^2\sigma^2} & 0 & \\ 0 & (1-p)J_{AA} & \\ \hline & & (1-p)\boldsymbol{J_{aa}} \end{array}\right]. \quad (5)$$

For the models b_1 and b_2 where $b(t_n)$ is deterministic the FIM is

$$\boldsymbol{J}_{b_i} = \left[\begin{array}{cc|c} J_{\sigma^2\sigma^2} & 0 & \\ 0 & J_{AA}^b & \\ \hline & & \boldsymbol{J_{aa}^b} \end{array}\right], \qquad i = 1, 2, \quad (6)$$

where $J_{AA}^b = \sum_{n=1}^{N} J_{AA}(t_n)I(b(t_n) = 1)$ and $\boldsymbol{J_{aa}^b} = \sum_{n=1}^{N} \boldsymbol{J_{aa}}(t_n)I(b(t_n) = 1)$.

Estimation of σ^2 is at best unaffected while the CRBs of A and $\boldsymbol{a}$ increase by at least a factor of $1/(1 - p)$ for model b_0 and by approximately the same factor for models b_1, b_2. In the latter case the exact value depends on A, $\boldsymbol{a}$, $b(t_n)$ and T. These CRBs represent the minimum variance attainable by any unbiased estimator when observations are missing.

3.1.2. Bias

Assuming 1) the location of the missing observations are known, 2) the SNR is sufficiently high and 3) T is chosen to avoid aliasing of the phase parameters [6], the MLE of $\boldsymbol{a}$ is approximately unbiased, this has been verified in simulations. However, a first order Taylor series expansion shows that even at high SNR the bias in the MLE of A is approximately $-pA$ and that in the MLE of σ^2 is approximately $A^2(1 - p)p$. Note that for $p = 0$ the bias is zero.

3.2. The DPT

The DPT utilises delay and conjugation operations to transform the PPS into a single tone, the DPT kernel. Frequency estimation techniques are then used to estimate the frequency of the DPT kernel $\mathrm{D}\mathcal{P}_m$, which is a simple function of a_m. For a LFM signal, one cycle of the DPT algorithm is required to estimate a_2, after which its effect is removed by demodulation, $z(t_n)\exp(-\jmath a_2 t_n^2)$. The demodulated signal is a single tone and the remaining parameters are found by MLE.

The effects of missing observations on the DPT can be assessed by substituting (1) into the $\mathrm{D}\mathcal{P}_2$ kernel yielding

$$\mathrm{D}\mathcal{P}_2 = b(t_n)b(t_n - \tau)A^2 s(t_n)s^*(t_n - \tau) + b(t_n - \tau)As^*(t_n - \tau)w(t_n) + b(t_n)As^*(t_n)w^*(t_n - \tau) + w(t_n)w^*(t_n - \tau), \quad (7)$$

where the optimal delay is $\tau = N/2$ [4]. The first term is the single tone while the other terms are considered noise. The expectation of the signal term at high SNR under model b_0 is $(1 - p)^2 A^2 s(t_n)s^*(t_n - \tau)$, so on average only a proportion $(1 - p)^2$ of the signal term remains, in effect reducing the sample size. This effect can be even more severe for blocks of missing observations, consider model b_1 with $p = 0.5$, then for the optimal $\tau = N/2$ the signal term is removed.

Results on the bias of the MLE carry over to the DPT. Again $\hat{a}_2$ appears unbiased at high SNR, although its variance is generally larger. This, coupled with a much smaller allowable range before aliasing occurs, means the SNR threshold increases and is larger than that of the MLE. The remaining parameters are found using MLE, hence the approximate biases found for $\hat{A}$ and $\widehat{\sigma^2}$ hold.

To summarise, the effect of missing observations on the DPT is an increase in the breakdown threshold SNR and an increase in variance even at high SNR.

4. PROPOSED METHODS

Following from the results of the previous section, modifications to the MLE and DPT estimator are suggested to improve their performance for missing observations based on the expectation-maximisation algorithm and varying the DPT delay parameter.

4.1. The Expectation-Maximisation Algorithm

As mentioned, the best performance is attained when the locations of the missing observations are known. This suggests estimating $b(t_n)$ so that noise only observations can be separated from those containing the signal. The likelihood of the signal parameters can then be maximised over locations where the signal is likely to exist.

This is essentially the nature of the Expectation-Maximisation (EM) algorithm [7, 8], as summarised in the context of LFM signal estimation in Table 1. Briefly, the algorithm consists of two steps. First, the expected likelihood function conditioned on the current estimates and the observations is found. Second, the expected likelihood is maximised with respect to the parameters, updating the estimates. The equations comprising the maximisation step are similar to the MLE except that $\hat{\mathrm{P}}^k(b(t_n) = 1|z(t_n))$, the conditional probability that the signal is present at t_n, weights the observations. This means more weight is placed on observations where the signal is likely to be present. The interpretation of the remaining conditional probabilities and the probability density functions f are apparent. Although the algorithm is iterative it is stable, convergence to at least a local maximum of the likelihood is certain.

Table 1. The EM algorithm for estimating the parameters of a PPS with missing observations.

1. **Initialisation:**
   Set k=0 and assume an initial value for $\hat{p}^k$.
   Obtain an initial estimate, $\hat{\boldsymbol{\theta}}^k$, using MLE.
2. **Expectation Step:** Compute the conditional probabilities

$$f(z(t_n);\hat{\boldsymbol{\theta}}^k) = f(z(t_n);\hat{\boldsymbol{\theta}}^k|b(t_n)=0)\hat{p}^k + f(z(t_n);\hat{\boldsymbol{\theta}}^k|b(t_n)=1)(1-\hat{p}^k)$$

$$\hat{\mathrm{P}}^k(b(t_n)=0|z(t_n)) = \frac{f(z(t_n);\hat{\boldsymbol{\theta}}^k|b(t_n)=0)\hat{p}^k}{f(z(t_n);\hat{\boldsymbol{\theta}}^k)}$$

$$\hat{\mathrm{P}}^k(b(t_n)=1|z(t_n)) = 1-\hat{\mathrm{P}}^k(b(t_n)=0|z(t_n))$$

3. **Maximisation Step:** Update the parameter estimates

$$\hat{\boldsymbol{a}}_0^{k+1} = \underset{\boldsymbol{a}}{\operatorname{argmax}} \left| \sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=1|z(t_n)) s_0^*(t_n) z(t_n) \right|$$

$$\hat{a}_0^{k+1} = \arg \sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=1|z(t_n)) \hat{s}_0^{*k+1}(t_n) z(t_n)$$

$$\hat{A}^{k+1} = \frac{\sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=1|z(t_n)) \Re\left\{\hat{s}_0^{*k+1}(t_n) z(t_n)\right\}}{\sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=1|z(t_n))}$$

$$\widehat{\sigma^2}^{k+1} = \frac{1}{N}\sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=0|z(t_n)) \, |z(t_n)|^2 + \hat{\mathrm{P}}^k(b(t_n)=1|z(t_n)) \left| z(t_n) - \hat{A}^{k+1}\hat{s}^{k+1}(t_n)\right|^2$$

$$\hat{p}^{k+1} = \frac{1}{N}\sum_{n=1}^{N} \hat{\mathrm{P}}^k(b(t_n)=0|z(t_n))$$

   where $\boldsymbol{a}_0 = (a_1, \ldots, a_M)^\mathsf{T}$, $s_0(t_n) = s(t_n)\exp(-\jmath a_0)$.
4. **Check for convergence:**
   If $|\hat{\theta}^{k+1} - \hat{\theta}^k| < \epsilon|\hat{\theta}^k|$ for each element of $\hat{\theta}$ stop, otherwise set $k \to k+1$ and go to step 2.

4.2. A Modified DPT Algorithm

Recall the example where if the first $0.5N$ observations are missing then due to the optimal delay parameter being $0.5N$ for LFM signals, the signal term is removed from the kernel. By altering τ to $0.2N$, $0.3N$ samples of the kernel will correspond to the signal term. Since the mean square error (MSE) in $\hat{a}_2$ is fairly constant over the range of delays $0.1N \leq \tau \leq 0.5N$ [4], there is little loss in optimising τ over this range. The criterion for optimality is chosen to be the empirical MSE of the residuals, $\widehat{\sigma^2}$, that is, minimise $\widehat{\sigma^2}(\tau_q)$ over several delays, $q = 1, \ldots, Q$. $\tau_q = 0.1N + 0.05(q-1)$, $Q = 8$, was used here.

The EM algorithm is also used to estimate the frequency of the $\mathrm{D}\mathcal{P}_2$ kernel and the demodulated signal. Again, error in frequency estimation is reduced by weighting samples where the signal is likely to exist more heavily. Table 2 summarises the procedure.

5. SIMULATIONS AND DISCUSSION

Simulations were performed for a LFM signal with $A = 1$, $\boldsymbol{a} = (\pi/100, \pi/2, 0)^\mathsf{T}$, $N = 31$ and $t_1 = -15$ over 100 independent Monte Carlo realisations for the DPT, MLE, modified DPT and the EM algorithm. Comparisons were made with the CRBs of (5) and (6) where appropriate for the three missing observations models b_0, b_1 and b_2 over $0 \leq p \leq 0.4$.

For the case of no missing observations it was found that all estimators achieved the CRB and were unbiased above the DPT and modified DPT SNR threshold of 5dB, although the MLE and EM algorithm perform well at even lower SNRs. Figure 1 shows the MSE of $\hat{a}_2$ for this case.

For even a small number of missing observations, $p = 0.05$, the MLE and DPT estimates for A and σ^2 became very poor.

At $p = 0.1$ the EM and modified DPT estimators perform well, the latter still achieving the CRB above 5dB, however, the MLE and DPT begin to show a tendency to be consistently above the CRB for $\hat{a}_2$ and $\hat{a}_1$, the effect being most apparent for model b_1. Figure 2 shows the MSE of $\hat{a}_1$ for model b_1.

For $p = 0.2$ the SNR threshold for the DPT increased to 15dB for model b_0, while the modified DPT maintains a threshold of 5dB, as shown in Figure 3 for $\hat{a}_1$. For models b_1 and b_2 the SNR threshold varies from 5 to 10dB with the phase parameters showing a deviation from the CRB for both the MLE and DPT. However, the EM and modified DPT still reach the CRB above 5dB.

The DPT essentially fails for $p = 0.4$ with a MSE far exceeding the CRB for all parameters. Even the MLE performs very poorly for models b_1 and b_2. However, the EM and modified DPT still attain the CRB with SNR thresholds ranging from 5dB to 15dB.

The phase estimates appear to be unbiased for all estimators operating above their SNR threshold, the same is true for the EM

Table 2. The modified DPT algorithm incorporating the EM algorithm and a search for the optimal DPT delay parameter.

1. **Initialisation:**
   Choose a set of delays, $0.1N \leq \tau_q \leq 0.5N$, $q = 1, \ldots, Q$ and set $q = 1$.
2. **Compute the residuals:**
   Use the DPT with delay τ_q to estimate $\hat{\boldsymbol{\theta}}_q$, $\widehat{\sigma^2}_q$.
   Estimate the parameters using the EM algorithm of Table 1 where in the first equation of Step 3 maximisation is carried out over frequency, a_1.
   If $q = Q$ go to step 3, otherwise set $q \leftarrow q + 1$ and go to step 2.
3. **Optimise the DPT delay parameter:**
   Choose $\hat{\boldsymbol{\theta}}_q$ such that $\widehat{\sigma^2}_q$ is minimised.

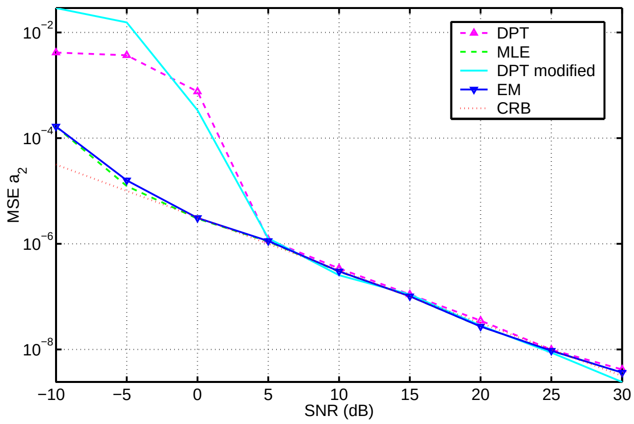

Fig. 1. MSE of $\hat{a}_2$ for $p = 0$.

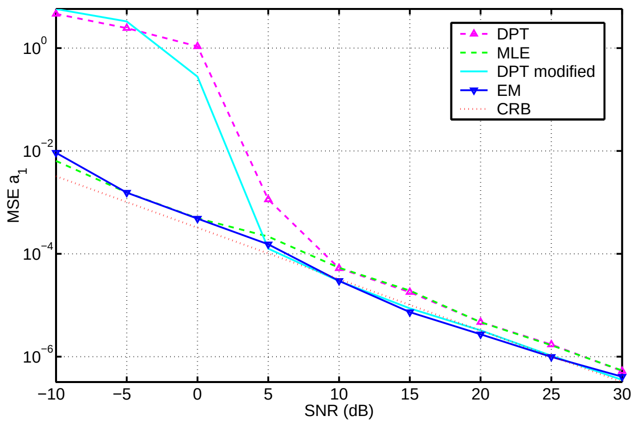

Fig. 2. MSE of $\hat{a}_1$ for $p = 0.1$ and model b_1.

and modified DPT estimates of A and σ^2. However, the MLE and DPT estimates of A and σ^2 were clearly biased in very close agreement with the Taylor series approximations.

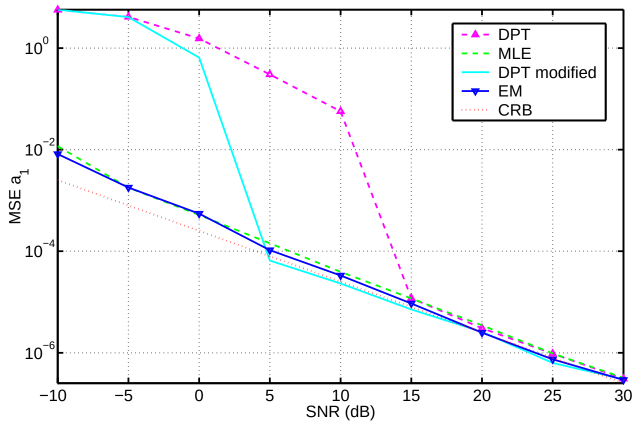

Fig. 3. MSE of $\hat{a}_1$ for $p = 0.2$ and model b_0.

6. CONCLUSION

The problem of estimating the parameters of PPS when observations are missing was considered. It was shown that for MLE the minimum CRB attainable is inversely proportional to the probability of a missed observation. It was also shown that the estimates of A and σ^2 are biased. The failure of the DPT was noted and explained, a modified DPT which made use of the EM algorithm and the freedom to optimise the delay parameter was proposed. The performance of the modified DPT and an EM formulation of the MLE estimator were shown to attain the CRB while existing methods failed as the number of missing observations increased.

7. REFERENCES

[1] M. Maksimov et al., *Radar Anti-Jamming Techniques*, Artech House, 1980.

[2] T. Abatzoglou, "Fast maximum likelihood joint estimation of frequency and frequency rate," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 6, pp. 708–715, November 1986.

[3] P. Djurić and S. Kay, "Parameter estimation of chirp signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 12, pp. 2118–26, December 1990.

[4] S. Peleg and B. Friedlander, "The discrete polynomial-phase transform," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 1901–14, August 1995.

[5] J. Legg and D. Gray, "Performance bounds for polynomial phase parameter estimation with nonuniform and random sampling schemes," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 331–7, February 2000.

[6] J. Ängeby, "Aliasing of polynomial-phase signal parameters," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1488–1491, May 2000.

[7] R. Redner and H. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, April 1984.

[8] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1996.