

# NON-LINEAR INSTANTANEOUS LEAST SQUARES AND ITS HIGH SNR ANALYSIS

J. ÄNGEBY, M. VIBERG AND T. GUSTAFSSON

Signal Processing Group, Department of Signals and Systems  
Chalmers University of Technology, Göteborg, Sweden

## ABSTRACT

A novel approach for signal parameter estimation, named the Non-Linear *Instantaneous* Least Squares (NILS) estimator, is proposed and a high SNR statistical analysis of the estimates is presented. The algorithm is generally applicable to deterministic signal in noise models. However, it is of particular interest in applications where the “conventional” non-linear least squares criterion suffers from numerous local minima. The key idea here is to apply a sliding window to estimate the instantaneous amplitude, which is then used in a separable least squares criterion-function. Hereby the radius of attraction of the global minimum is under the control of the user, which makes the NILS approach advantageous to use in practical applications. Numerical results using polynomial-phase signals validate the theoretical results.

## 1. INTRODUCTION

Deterministic signal models are important and widely used in applications, and the estimation of such signal parameters has been studied extensively. For the case of wide sense stationary signals, there is a large number of estimation approaches that can be used and there is a rich literature on the estimation of such signal parameters, see for example [5]. However, if the signal is non-stationary, or if it has been non-uniformly sampled, then there are not many estimation techniques available. In theory, the Non-Linear Least Squares (NLLS) approach to signal parameter estimation can be used on such signals. However, for many signal models the NLLS approach suffers from severe numerical problems. Some methods have been developed that can estimate the parameters of certain classes of non-stationary signals. For example, motivated by applications such as radar and mobile tele-communication, the estimation of Polynomial-Phase Signals (PPS) parameters has received considerable attention recently [3, 4, 6]. Here, a novel approach to signal parameter estimation named the Non-Linear *Instantaneous* Least Squares (NILS) estimator is presented. The key idea is to use a sliding window to es-

timate the instantaneous amplitude, which is then used in a separable least squares criterion-function.

## 2. MOTIVATION FOR THE NILS APPROACH

Consider the following deterministic signal model:

$$y_k = b s_k(\boldsymbol{\theta}) + e_k, \quad k = 1, \dots, N, \quad (1)$$

where  $\boldsymbol{\theta}$  denotes the signal parameters,  $e_k$  the noise,  $b$  a complex-valued amplitude and  $s_k(\boldsymbol{\theta})$  is the deterministic signal waveform. For simplicity, it is assumed that the noise is independent and identically distributed with variance  $\sigma^2$ . Modification of the analysis to the general case is straightforward. A well-known approach to estimate the signal parameters is to use a NLLS fit of  $b s_k(\boldsymbol{\theta})$  to  $y_k$  with respect to  $b$  and  $\boldsymbol{\theta}$ :

$$\{\hat{b}, \hat{\boldsymbol{\theta}}\} = \arg \min_{b, \boldsymbol{\theta}} \sum_{k=1}^N |y_k - b s_k(\boldsymbol{\theta})|^2. \quad (2)$$

Since the NLLS criterion is linear in  $b$ , it can be concentrated w.r.t.  $\boldsymbol{\theta}$ . Expressing the estimate of  $b$  as a function of  $\boldsymbol{\theta}$  leads to

$$\hat{b}(\boldsymbol{\theta}) = \frac{\sum_{k=1}^N y_k s_k^*(\boldsymbol{\theta})}{\sum_{k=1}^N |s_k(\boldsymbol{\theta})|^2} = \frac{\mathbf{s}_N^H(\boldsymbol{\theta}) \mathbf{y}_N}{|\mathbf{s}_N(\boldsymbol{\theta})|^2}, \quad (3)$$

where  $\mathbf{y}_N = (y(1), \dots, y(N))^T$ ,  $(\cdot)^*$  denotes complex conjugation,  $(\cdot)^H$  is complex conjugate transpose and  $\mathbf{s}_N(\boldsymbol{\theta})$  is defined conformably with  $\mathbf{y}_N$ . Using (3) in (2) gives

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \sum_{k=1}^N \left| y_k - \frac{\mathbf{s}_N^H(\boldsymbol{\theta}) \mathbf{y}_N}{|\mathbf{s}_N(\boldsymbol{\theta})|^2} s_k(\boldsymbol{\theta}) \right|^2. \quad (4)$$

In the case of Gaussian noise, this approach is equivalent to the Maximum Likelihood (ML) estimator, which is well known to be statistically efficient. However, for many signal models of interest, NLLS suffers from severe numerical problems which makes it difficult to find the global minimum. Firstly, the Radius of Attraction (RoA) of the global minimum, defined as the maximum distance from the minimum that a gradient-based search can be originated and still find it, may be very small (in the case of a sinusoidal signal,

---

THIS WORK WAS SUPPORTED IN PART BY ERICSSON MICROWAVE SYSTEMS AB AND ERICSSON INFOTECH AB.

Telephone:+46 31 772 1792, fax: +46 31 772 1782, e-mail: tonyg@s2.chalmers.se

the RoA is proportional to  $1/N$  [7]). Secondly, in nearly ambiguous situations, a local minimum can yield almost the same value of the noise-less criterion as does the global minimum. In the presence of noise, the local minimum can then erroneously become global. Due to the problems inherent in the NLLS approach, it is only on rare occasions that it can be used.

### 3. THE NILS APPROACH

To overcome the mentioned shortcomings in the NLLS approach, a ‘‘smoothing’’ in the criterion is proposed. The idea is to replace the NLLS estimate of the amplitude  $b$  with a time and signal-parameter dependent local LS estimate, based on observations from some sliding window. In the following, this instantaneous amplitude estimate will be denoted  $\hat{b}_k(\boldsymbol{\theta})$ . For example, let  $\hat{b}_k(\boldsymbol{\theta})$  be a local amplitude estimate based on the data  $\{y_{k+l}\}_{l=0}^n$ . Here,  $n$  is a user-defined integer that determines the length of the sliding window (the length is  $n + 1$ ). For the sake of simplicity, a rectangular window will be used in what follows, although other windows (Hamming, Hanning, etc.) may be preferable. The least squares estimate of the local amplitude is then

$$\hat{b}_k(\boldsymbol{\theta}) = \frac{\mathbf{s}^H(k; \boldsymbol{\theta})\mathbf{y}(k)}{|\mathbf{s}(k; \boldsymbol{\theta})|^2}, \quad (5)$$

where

$$\mathbf{s}(k; \boldsymbol{\theta}) \stackrel{\text{def}}{=} (s_k(\boldsymbol{\theta}), \dots, s_{k+n}(\boldsymbol{\theta}))^T, \quad (6)$$

$$\mathbf{y}(k) \stackrel{\text{def}}{=} (y_k, \dots, y_{k+n})^T. \quad (7)$$

An additional smoothing in the NLLS criterion can be introduced by regularizing the amplitude estimate. The main reason for this is to avoid problems when  $|s_k(\boldsymbol{\theta})| \simeq 0$ . However, this option will not be used here, and the reader is referred to [2] for details. Replacing  $\hat{b}(\boldsymbol{\theta})$  in (4) with the instantaneous time-varying amplitude estimate  $\hat{b}_k(\boldsymbol{\theta})$ , gives the NILS criterion<sup>1</sup>

$$V_N(\boldsymbol{\theta}) = \sum_{k=1}^{N-n} \left| y_k \Leftrightarrow \frac{\mathbf{s}^H(k; \boldsymbol{\theta})\mathbf{y}(k)}{|\mathbf{s}(k; \boldsymbol{\theta})|^2} \mathbf{s}_k(\boldsymbol{\theta}) \right|^2. \quad (8)$$

The signal parameter estimates are the minimizing arguments of  $V_N(\boldsymbol{\theta})$ . The parameter estimates are found using a non-linear search that can be implemented on- or off-line. Consider the criterion (8). Due to the windowing, the parameter estimates are based on  $N \Leftrightarrow n$  fits of  $\hat{b}_k(\boldsymbol{\theta})s_k(\boldsymbol{\theta})$  to  $y_k$ . The  $n$  last observations are only used implicitly through

<sup>1</sup>Note that the criterion-function is closely related to the structured AR criterion introduced in [1], which establishes a link between NILS and prediction based estimators.

$\hat{b}_k(\boldsymbol{\theta})$ . Criterion (8) can be modified to overcome this drawback as follows:

$$V_N(\boldsymbol{\theta}) = \sum_{k=1}^{N-n} \left| \mathbf{y}(k) \Leftrightarrow \hat{b}_k(\boldsymbol{\theta})\mathbf{s}(k; \boldsymbol{\theta}) \right|^2, \quad (9)$$

where  $\hat{b}_k(\boldsymbol{\theta})$ ,  $\mathbf{s}(k; \boldsymbol{\theta})$  and  $\mathbf{y}(k)$  are defined in (5)-(7). The modified criterion is the mean of  $N \Leftrightarrow n$  consecutive NLLS criterion-functions, based on the observations in  $\mathbf{y}(k)$ ,  $k = 1, \dots, N \Leftrightarrow n$ . Using (5) in (9) implies

$$V_N(\boldsymbol{\theta}) = \sum_{k=1}^{N-n} \left| \boldsymbol{\Pi}^\perp(k; \boldsymbol{\theta})\mathbf{y}(k) \right|^2, \quad (10)$$

where

$$\boldsymbol{\Pi}^\perp(k; \boldsymbol{\theta}) \stackrel{\text{def}}{=} \mathbf{I} \Leftrightarrow \boldsymbol{\Pi}(k; \boldsymbol{\theta}), \quad (11)$$

$$\boldsymbol{\Pi}(k; \boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{1}{|\mathbf{s}(k; \boldsymbol{\theta})|^2} \mathbf{s}(k; \boldsymbol{\theta})\mathbf{s}^H(k; \boldsymbol{\theta}). \quad (12)$$

Note that  $\boldsymbol{\Pi}(k; \boldsymbol{\theta})$  projects onto the space spanned by  $\mathbf{s}(k; \boldsymbol{\theta})$ , which is a one-dimensional instantaneous signal subspace. Hereby a link is established between NILS and estimators based on signal subspaces. Henceforth, the NILS criterion (9) will be referred to as subspace-based, and the criterion (8) will be referred to as prediction-based.

Generally, the subspace-based NILS approach is to be preferred over the prediction-based. The subspace-based criterion function has empirically found to be ‘‘smoother’’ in the sense that there are fewer local minima. Still, the global minimum of the subspace-based criterion function is more ‘‘pointed’’, which gives it a slightly better statistical performance, than the prediction-based approach. An appealing property is also that the subspace-based NILS estimator is equivalent to NLLS if the window length is chosen to  $n + 1 = N$ . For the above reasons, the subspace-based criterion (9) will be used henceforth. The following example highlights the basic advantages of the NILS approach.

**Example 1** Consider a complex valued sinusoid, with angular frequency  $\omega_0 = 1.5$ . In Figure 1, the mean value of the criterion (10) is plotted for increasing  $n$ . The RoA of the global minimum in the NILS criterion depends on the window length. As  $n$  increases, the RoA decreases and the minimum becomes more pointed. Note that for  $n = 1$ , there are no local minima.

Seen from a numerical/computational point of view, a small NILS window length is preferred, since a gradient-based search is then more likely to converge to the global minimum. However, it is intuitively obvious that the statistical performance of the NILS estimates is improved (i.e. the MSE decreases) as  $n$  increases. These properties can be utilized in a search strategy as follows: Initiate the search using a short window to assure convergence to the global minimum. Then, as the search converges, the window length can be increased in order to improve the accuracy.

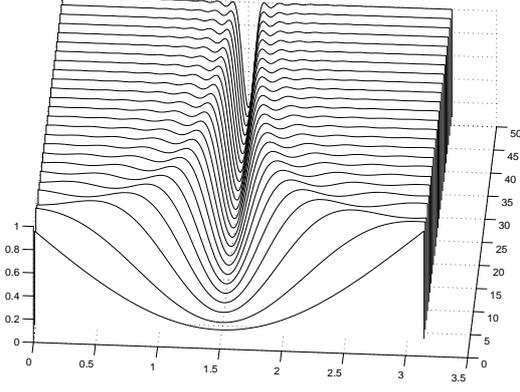


Figure 1: Mean value of the criterion-function (10) as a function of  $n$ . Single cisoid,  $\omega_0 = 1.5$ , SNR=10 dB.

We conclude this section by mentioning that the NILS approach can be easily extended to include general windows, multi-component signal models, and also in fact to stochastic signal models [2].

#### 4. STATISTICAL ANALYSIS

Assessing the statistical performance of an estimator is in general an exceedingly complex task. However, under simplifying assumptions, it is possible to analytically derive results on the statistical distribution of the estimates. Here, a first-order expression for the covariance matrix of the NILS parameter estimates is presented, which is valid for high enough SNR. Here, the SNR is defined as  $\text{SNR} = |b|^2/\sigma^2$ , which is a natural definition only if the model signal  $s_k(\boldsymbol{\theta})$  is of constant-modulus equal to one. For simplicity, we let  $|b|$  be fixed and assume that  $\sigma \rightarrow 0$ . The analysis is applicable to any smooth and uniquely identifiable signal model. The identifiability condition assumed to hold is

$$bs_k(\boldsymbol{\theta}) = \bar{b}s_k(\bar{\boldsymbol{\theta}}), k = 1, \dots, N \Leftrightarrow \boldsymbol{\theta} = \bar{\boldsymbol{\theta}}, b = \bar{b}. \quad (13)$$

To begin with, the consistency of the estimates is established. Under the stated assumptions, the criterion function  $V_N(\boldsymbol{\theta})$  (10) converges, uniformly in  $\boldsymbol{\theta}$ , to the limiting function  $\bar{V}_N(\boldsymbol{\theta})$  as  $\sigma \rightarrow 0$ . Here,

$$\bar{V}_N(\boldsymbol{\theta}) = |b_0|^2 \sum_{k=1}^{N-n} |\boldsymbol{\Pi}^\perp(k; \boldsymbol{\theta}) \mathbf{s}(k; \boldsymbol{\theta}_0)|^2, \quad (14)$$

where  $b_0$  and  $\mathbf{s}(k; \boldsymbol{\theta}_0)$  denote the true values. Clearly,  $\bar{V}_N(\boldsymbol{\theta}) \geq 0$ , with equality if and only if  $\mathbf{s}(k; \boldsymbol{\theta}_0) = c_k \mathbf{s}(k; \boldsymbol{\theta})$  for all  $k$ , and for some set of scalars  $c_k$ . However, in view of (13) this is possible only for  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  and we have proven the following:

**Theorem 1** *Let the signal parameters  $\boldsymbol{\theta}$  be defined on a compact set  $\Theta$ , and assume that the true value  $\boldsymbol{\theta}_0$  is an inner point of  $\Theta$ . Further, assume that the gradient of  $s_k(\boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$  is bounded on  $\Theta$  and that (13) holds. Then we have*

$$\hat{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}_0 \text{ as } \sigma \rightarrow 0. \quad (15)$$

Next, the attention is turned to the variance of the estimates. Since  $\hat{\boldsymbol{\theta}}$  minimizes  $V_N(\boldsymbol{\theta})$ , we have  $V'_N(\boldsymbol{\theta}) = 0$  at  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , where  $V'_N(\boldsymbol{\theta})$  denotes the gradient. For high SNR, a first-order Taylor expansion yields

$$0 = V'_N(\boldsymbol{\theta}_0) + V''_N(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} \Leftrightarrow \boldsymbol{\theta}_0) + o_p(|V'_N(\boldsymbol{\theta}_0)|), \quad (16)$$

where  $V''_N(\boldsymbol{\theta}_0)$  denotes the Hessian and  $o_p(\cdot)$  is order in probability. Defining

$$\mathbf{H} = \lim_{\sigma \rightarrow 0} V''_N(\boldsymbol{\theta}_0) \quad (17)$$

it now follows

$$\hat{\boldsymbol{\theta}} \Leftrightarrow \boldsymbol{\theta}_0 = \Leftrightarrow \mathbf{H}^{-1} V'_N(\boldsymbol{\theta}_0) + o_p(|V'_N(\boldsymbol{\theta}_0)|), \quad (18)$$

provided  $\mathbf{H} > 0$ . Further, let

$$\mathbf{Q} = \lim_{\sigma \rightarrow 0} \frac{1}{\sigma^2} \mathbb{E}[V'_N(\boldsymbol{\theta}_0) V'^T_N(\boldsymbol{\theta}_0)]. \quad (19)$$

For high SNR, the mean square error (MSE) matrix of the estimation error is then given by

$$\mathbb{E}[(\hat{\boldsymbol{\theta}} \Leftrightarrow \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} \Leftrightarrow \boldsymbol{\theta}_0)^T] = \sigma^2 \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1} + o(\sigma^2). \quad (20)$$

Evaluation of  $\mathbf{H}$  and  $\mathbf{Q}$  leads to the following:

**Theorem 2** *Let the conditions of Theorem 1 hold, and assume in addition that  $s_k(\boldsymbol{\theta})$  has bounded derivatives up to order three. Further, assume that the asymptotic Hessian matrix  $\mathbf{H}$  is positive definite. Then the MSE matrix of the estimation error is given by (20), where*

$$\mathbf{H} = 2|b_0|^2 \sum_{k=1}^{N-n} \text{Re} \left\{ \mathbf{G}^H(k) \boldsymbol{\Pi}^\perp(k; \boldsymbol{\theta}) \mathbf{G}(k) \right\} \quad (21)$$

$$\mathbf{Q} = 2|b_0|^2 \sum_{k=1}^{N-n} \sum_{l=\max\{1, k-n\}}^{\min\{N-n, k+n\}} \text{Re} \{ \mathbf{Z}(k, l) \}. \quad (22)$$

Here,  $\mathbf{Z}(k, l) = \mathbf{G}^H(k) \boldsymbol{\Pi}^\perp(k; \boldsymbol{\theta}) \mathbf{J}_{k-l} \boldsymbol{\Pi}^\perp(l; \boldsymbol{\theta}) \mathbf{G}(l)$ ,  $\mathbf{G}(k) = \text{d}\mathbf{s}(k; \boldsymbol{\theta})/\text{d}\boldsymbol{\theta}$  denotes the Jacobian matrix, and  $\mathbf{J}_k$  is an  $(n+1) \times (n+1)$  matrix of zeros, except at the  $k$ :th super-diagonal, which is all ones (e.g.  $\mathbf{J}_n$  has a single one in the upper right corner).

The proof consists of straightforward calculations, and the details will be published elsewhere. As a corollary, an expression for the Cramér-Rao lower bound assuming Gaussian noise is obtained. This is due to the fact that for  $n =$

$N \Leftrightarrow 1$ , the NILS estimator coincides with the ML estimator (NLLS), which is known to be statistically efficient at high SNR and/or large  $N$ . For  $n = N \Leftrightarrow 1$  we have  $\mathbf{H} = \mathbf{Q}$ , and the CRB inequality is expressed as

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \geq \frac{2}{\text{SNR}} \left[ \text{Re} \left\{ \mathbf{G}^H \boldsymbol{\Pi}^\perp(\boldsymbol{\theta}) \mathbf{G} \right\} \right]^{-1}, \quad (23)$$

where, for  $n = N \Leftrightarrow 1$ , the  $k$ -argument of  $\mathbf{G}$  has been omitted.

## 5. POLYNOMIAL PHASE SIGNAL EXAMPLE

The analytical expression for the high SNR covariance matrix of the estimation error is rather involved. Thus, one has to resort to numerical comparisons to get further insight into the performance in realistic scenarios. Since NILS coincides with NLLS for  $n = N \Leftrightarrow 1$ , it is of particular interest to investigate how close to the NLLS performance one comes for smaller values of  $n$ . Here, the results of applying the NILS estimator to a 4:th order polynomial phase signal (PPS) are presented. The signal model is thus

$$s_k(\boldsymbol{\theta}) = b e^{j\phi(\boldsymbol{\theta};k)} + e_k \quad (24)$$

where the true parameter values are  $b_0 = 1$  and

$$\phi(\boldsymbol{\theta}_0; k) = 2\pi \sum_{l=1}^4 \theta_l k^l, \quad \theta_l = 4/N^l. \quad (25)$$

The noise  $e_k$  is a zero-mean white Gaussian random process with variance equal to  $1/\text{SNR}$ . The number of samples is  $N = 200$  and the SNR is varied from -9 dB to 30 dB. The NILS estimator is applied with  $n = 1$  and  $n = 9$  respectively. Figure 2 shows the theoretical and empirical RMS errors for  $\hat{\theta}_4$ . The theoretical curves are based on Theorem 2, whereas the empirical results are obtained from 100 independent Monte-Carlo runs for each displayed SNR-value. The corresponding curves for the other signal parameters show the same characteristic behaviour, except that the magnitude of the RMSE decreases drastically with increasing coefficient number (i.e.  $\theta_4$  is more accurately estimated than  $\theta_3$ , etc). However, the relative efficiency, defined as  $\text{RMSE}/\text{CRLB}$ , is approximately the same for all PPS parameters and lies in the range 1 (statistical efficiency) to 3-4, depending on the length of the sliding window.

Based on the example presented here and other similar experiments we can draw the following conclusions:

- The empirical RMS errors closely follow the theoretical values down to an SNR threshold.
- The SNR threshold is essentially the same for all PPS parameters, and it decreases with increasing window length.
- The NILS estimates are close to statistical efficiency already at small values of  $n$  ( $N/n \approx 10$ ). The ratio of the estimation error variance to the CRB is approximately the same for all PPS parameters.

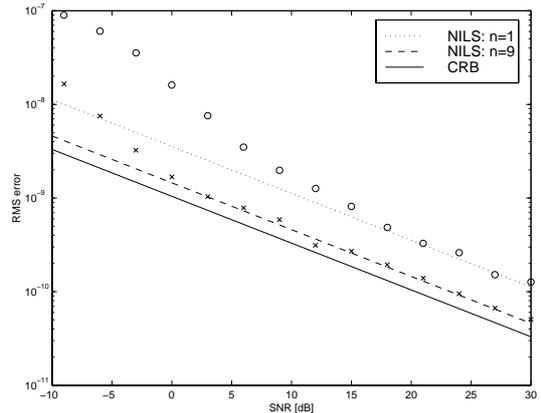


Figure 2: Theoretical (dotted and dashed line) and empirical ('o' and 'x') RMS errors for  $\hat{\theta}_4$  versus the signal-to-noise ratio. The NILS estimator is applied with  $n = 1$  and  $n = 9$  respectively. The solid line is the Cramér-Rao lower bound.

The second conclusion deserves to be stressed, since most known suboptimal PPS estimators have a rather high SNR threshold for polynomial-phase signals of high order [2]. The advantages compared to NLLS and to the previously known suboptimal methods should make NILS a viable alternative in practical applications.

## 6. REFERENCES

- [1] J. Ängeby. "Structured Auto-Regressive Instantaneous Phase and Frequency Estimation". In *ICASSP*, volume 3, pages 1768–1771. IEEE, May 1995.
- [2] J. Ängeby. "Estimating Signal Parameters using the Instantaneous Non-Linear Least Squares Approach". *Submitted to IEEE Sig. Proc.*, 1998.
- [3] S. Barbarossa, A. Scaglione, and G. B. Giannakis. "Product High-Order Ambiguity Function for Multi-component Polynomial-Phase Signal Modeling". *IEEE Trans. Sig. Proc.*, 46(3):691–708, March 1998.
- [4] P. M. Djuric and M. Kay. "Parameter Estimation of Chirp Signals". *IEEE Trans. Sig. Proc.*, 38:2118–2126, Dec 1990.
- [5] S. M. Kay. *Fundamentals of Statistical Signal Processing*. Signal Processing. Prentice Hall, 1993.
- [6] S. Peleg and B. Porat. "Estimation and Classification of Polynomial-Phase Signals". *IEEE Trans. Inform. Theory*, 37:422–430, March 1991.
- [7] P. Stoica, R. Moses, B. Friedlander, and T. Söderström. "Maximum Likelihood Estimation of the Parameters of Multiple Sinusoids From Noisy Measurements". *IEEE Trans. ASSP*, ASSP-37:378–392, 1989.