

Maximum Likelihood Estimation of Wiener Models

Anna Hagenblad and Lennart Ljung
 Division of Automatic Control
 Department of Electrical Engineering
 Linköpings universitet
 SE-581 83 Linköping, Sweden

email: annah@isy.liu.se, ljung@isy.liu.se

Abstract

A Wiener model consists of a linear dynamic system followed by a static nonlinearity. The input and output are measured, but not the intermediate signal. We discuss the Maximum Likelihood estimate for Gaussian measurement and process noise, and the special cases when one of the noise sources is zero.

1 The Wiener Model

The Wiener model is depicted in Figure 1. It consists

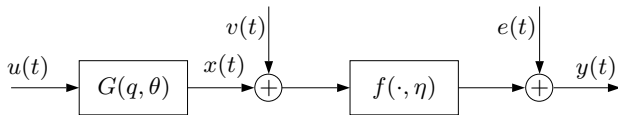


Figure 1: The Wiener model. $u(t)$ denotes the input, $y(t)$ the output. The intermediate signal $x(t)$ is not measurable. $v(t)$ denotes process noise and $e(t)$ measurement noise. The linear dynamic subsystem is denoted $G(q, \theta)$ and the static nonlinear subsystem $f(\cdot, \eta)$. q denotes the time shift operator, θ the parameters of the linear system and η the parameters of the nonlinear system.

of a linear dynamic system G in series with a static nonlinearity f . We will consider two noise sources: The measurement noise $e(t)$ and the process noise $v(t)$.

The output of the Wiener model is:

$$y(t) = f(G(q, \theta)u(t) + v(t), \eta) + e(t) \quad (1)$$

Identification of Wiener models has been treated in several papers. [6], [7] and [1] consider the case with measurement noise only. If the input signal is Gaussian, Busgang's theorem [2] can be used to show that linear identification methods will give consistent estimates despite the nonlinearity. A prediction error criterion may also be useful in this case. A problem is that the criterion may have several local minima, so to find the minimum from a numerical search, a good initialization is needed. [3] suggests that the method proposed in [5] is used as an initial estimate.

If instead the measurement noise is zero, but there is process noise, it may be more interesting to minimize the error between $G(q, \theta)u(t)$ and $f^{-1}(y(t), \eta)$. [8] considers this case.

Few papers deal with the case when there are both process noise and measurement noise. In this case, the prediction error method (PEM) is less attractive since the predictor is hard to calculate.

2 The Maximum Likelihood Estimate

The Maximum Likelihood (ML) estimate is defined as the one maximizing the likelihood of the actual observations, $p_y(\theta, \eta; y_N^*)$. p_y here denotes the probability density function of y , where the observed outputs y_N^* are inserted. Using the intermediate signal $x(t)$ (cf. Figure 1) as a nuisance parameter, we have

$$\begin{aligned} p_y(\theta, \eta; y_N^*) &= \int_{\mathbf{R}^N} p_{x,y}(\theta, \eta, y_N^*) dx \\ &= \int_{\mathbf{R}^N} p_{y|x}(\theta, \eta; y_N^*) p_x(\theta, \eta; y_N^*) dx \\ &= \int_{\mathbf{R}^N} p_e(y(t) - f(x(t), \eta), \theta, \eta; y_N^*) \\ &\quad \cdot p_v(x(t) - G(q, \theta)u(t), \theta, \eta; y_N^*) dx \quad (2) \end{aligned}$$

Assuming that the process noise $v(t)$ and the measurement noise $e(t)$ are white and Gaussian, with zero mean and variance λ_v and λ_e , respectively, the above equation is equal to

$$p_y(\theta, \eta; y_N^*) = \left(\frac{1}{2\pi\sqrt{\lambda_e\lambda_v}} \right)^N \prod_{t=1}^N \int_{-\infty}^{\infty} e^{-\frac{1}{2}\epsilon(t)} dx(t) \quad (3)$$

where

$$\epsilon(t) = \frac{1}{\lambda_e} (y(t) - f(x(t), \eta))^2 + \frac{1}{\lambda_v} (x(t) - G(q, \theta)u(t))^2$$

Given θ , η , and data $\{u_N, y_N\}$, p_y can be calculated. Since each integral only depends on $x(t)$ for one particular t , they can be calculated in parallel, so there is no curse of dimensionality here. By a numerical search, we can thus maximize $p_y(\theta, \eta; y_N^*)$ and find the ML estimate.

For the special cases that either the process noise or the measurement noise is equal to zero, the criterion simplifies considerably. If the process noise $v(t)$ is zero, the likelihood function is simply

$$\begin{aligned} p_y(\theta, \eta; y_N^*) &= p_e(y(t) - f(x(t), \eta), \theta, \eta; y_N^*) \\ &= \prod_{t=1}^N \frac{1}{\sqrt{2\pi\lambda_e}} e^{-\frac{1}{2\lambda_e}(y(t) - f(x(t), \eta))^2} \end{aligned} \quad (4)$$

Maximizing this equation is equivalent to minimizing the criterion

$$V_N(\theta, \eta) = \frac{1}{N} \sum_{t=1}^N (y(t) - f(x(t), \eta))^2 \quad (5)$$

which is the prediction error criterion we recognize from, e.g., [7], [6] and [3].

If the measurement noise is zero, but we have process noise, and we also assume that the nonlinearity is invertible, we have

$$\begin{aligned} p_y(\theta, \eta; y_N^*) &= p_v(f^{-1}(y(t)) - G(q, \theta)u(t), \theta, \eta; y_N^*) \\ &= \prod_{t=1}^N \frac{1}{\sqrt{2\pi\lambda_v}} e^{-\frac{1}{2\lambda_v}(f^{-1}(y(t)) - G(q, \theta)u(t))^2} \end{aligned} \quad (6)$$

An equivalent criterion is then

$$V_N(\theta, \eta) = \frac{1}{N} \sum_{t=1}^N (f^{-1}(y(t)) - G(q, \theta)u(t))^2 \quad (7)$$

as used in [5] and [8].

3 Simulation Example

We will use the following very simple example:

$$x(t) = u(t-1) + v(t) \quad (8)$$

$$y(t) = ax^2(t) + e(t) \quad (9)$$

where the input u is white Gaussian noise with variance 1, the process noise v is also Gaussian with variance 4, and the measurement noise e is Gaussian with variance 1. These three signals are mutually independent. In [4], it is shown that if the process noise variance is large, the approximate prediction error criterion (5) will yield a biased estimate for this example system.

1000 data points were generated, using the parameter $a = 1$. The parameter value was then estimated from data using two different criteria, the ML criterion (3), and the approximate prediction error criterion (5). Both searches were initialized using the true parameter.

In a Montecarlo simulation, 1000 data sets as above were generated and used for estimation. The results

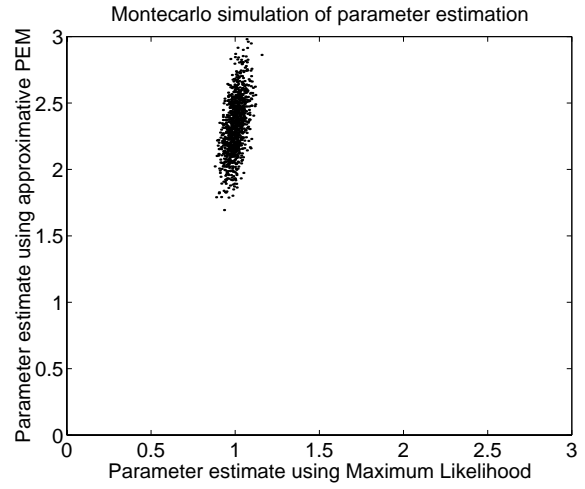


Figure 2: Montecarlo simulation. The estimate obtained from the approximative PEM is plotted against the value obtained from the ML estimate. The true value is $a_0 = 1$.

are shown in Figure 2. The ML estimates are centered around the true value $a = 1$, and also has a smaller variance than the approximate PEM, which is heavily biased around 2.4.

References

- [1] S. A. Billings and S. Y. Fakhouri. Identification of nonlinear systems using the Wiener model. *Electronics Letters*, 13(17):502–504, August 1977.
- [2] J. J. Bussgang. Crosscorrelation functions of amplitude-distorted Gaussian signals. Technical Report 216, MIT Research Laboratory of Electronics, 1952.
- [3] Anna Hagenblad. Aspects of the identification of wiener models. Technical Report Licentiate Thesis no. 793, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Nov 1999.
- [4] Anna Hagenblad. Inconsistency of an approximate prediction error method for Wiener model identification. Technical Report LiTH-ISY-R-2275, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, Sep 2000.
- [5] A. D. Kalafatis, L. Wang, and W. R. Cluett. Identification of Wiener-type nonlinear systems in a noisy environment. *International Journal of Control*, 66(6):923–941, 1997.
- [6] David Westwick and Michel Verhaegen. Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, 52:235–258, 1996.
- [7] Torbjörn Wigren. Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, 29(4):1011–1025, 1993.
- [8] Yucai Zhu. Parametric Wiener model identification for control. In *14th World Congress of IFAC*, pages 37–42, Beijing, China, July 1999.