Experiment design for the identification of a simple Wiener system

Michel Gevers, Matthias Caenepeel and Johan Schoukens

Abstract— **Results on experiment design for the identification of nonlinear systems are extremely scarce. This paper examines the identification and optimal input design for a very simple nonlinear system: a Wiener system composed of a Finite Impulse Response (FIR) system followed by a power nonlinearity: $(.)^n$. We first show that an expanding power ($n > 1$) increases the information about the estimated parameters, while a compressing power ($0 < n < 1$) decreases the information. We then formulate a simple optimal input design problem for the considered class of Wiener systems and show that solutions can be computed by restricting the class of considered input signals. We provide a solution which offers some intuitive insights for the case of a FIR(2) system with a square nonlinearity and where the inputs are restricted to be Gaussian.**

I. INTRODUCTION

Optimal experiment design for the identification of linear time-invariant systems is now a very mature field and solutions abound for a range of design criteria. The first developments date back to the 1970's with the pioneering work of Aoki and Staley, Mehra, Goodwin, Payne, Zarrop, Ng and Söderström. An excellent survey of these early results can be found in [5]. Most of the work of that period focused on open loop identification, and the optimality criteria that were minimized consisted in various functions of the asymptotic per sample parameter covariance matrix.

The activity in optimal experiment design took a new turn in the last decade, with the adoption of a wide range of application-oriented optimality criteria and the solution of such optimal design problems in both an open loop and a closed loop identification setup. A survey of these recent results can be found in [1].

Up until very recently, all results on optimal experiment design for system identification have addressed the identification of linear time-invariant systems. The first few results on the identification of nonlinear systems that we are aware of are [2], [7], [4], [3].

In [2] the authors explain that in the presence of nonlinearities the optimal inputs typically depend on the complete distribution of the input sequence. However, they isolate a number of simple situations where the optimal input only

M. Gevers is with ICTEAM, Université catholique de Louvain, B1348 Louvain la Neuve, Belgium, and with Dept ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B1050 Brussels, Belgium Michel.Gevers@uclouvain.be

M. Caenepeel and J. Schoukens are with Dept ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B1050 Brussels, Belgium mcm_caenepeel@hotmail.com, Johan.Schoukens@vub.ac.be

depends on a few higher order moments or where only deterministic inputs are considered, in which case the optimal design problem can often be reduced to the solution of a minimization problem involving polynomial matrix inequalities for which relaxation methods have recently been developed.

In [7] an experiment design problem is considered for the identification of the static nonlinearities of block structured nonlinear systems made up of the interconnection of known linear dynamic systems and unknown static nonlinearities. The nonlinear blocks are estimated as pointwise functions, and an input sequence is computed that minimizes an upper bound on the variance of these pointwise estimates.

In [4] an experiment design for the closed-loop identification of a class of nonlinear systems is considered, under a priori known bounds on the parameters and on model uncertainties. The problem is reformulated as a time-domain constrained optimal control problem over a finite time interval, where the control objective is a trajectory following problem. The optimal trajectory is itself obtained as the solution of a worst case norm of the parameter estimation error, where the parameter vector is computed by Least Squares estimation.

The paper [3] considers optimal input design for nonlinear FIR-type systems. It illustrates that the computational complexity becomes untractable when the number of lags in the nonlinear FIR model grows because the number of optimization variables becomes very large. The authors propose a suboptimal solution by restricting the inputs to take only a finite number of levels; in particular they present an example where the inputs are limited to three possible levels.

The purpose of the present paper is to add insights into the difficult and largely open field of optimal input design for nonlinear systems. We focus on a special class of such systems, namely a Wiener system whose linear dynamic part is an FIR(m) system and whose static nonlinear part raises the output of the FIR system to the power n. Our contribution is twofold. First we examine the effect of the nonlinearity on the information matrix, and hence the asymptotic parameter covariance matrix. We show in particular that an expanding nonlinearity, i.e. $n > 1$, actually improves the information about the estimated parameters, even if the signal to noise ratio at the output is kept constant as n varies. We then discuss the difficulties associated with the construction of optimal inputs for such Wiener systems, which were already highlighted in [3]. We propose a suboptimal solution, different from that adopted in [3], by restricting the inputs to be Gaussian. We present a complete solution for this restricted class of Gaussian inputs for the case of a FIR(2) system and

a square nonlinearity, and we discuss its possible extensions to FIR systems of higher order than $m = 2$ and to powers higher than $n = 2$.

The paper is organized as follows. We state the problem in Section II. In Section III we establish the relation between the order m of the FIR(m) system, the power n of the nonlinearity and the information matrix, and we show that an expanding nonlinearity improves the information. We illustrate these connections in Section IV for a static system, i.e. FIR(1). In Section V we study the case of an FIR(2) system followed by a square nonlinearity, and in Section VI we provide a suboptimal solution for such case by restricting to Gaussian inputs. Section VII presents simulations which validate the results of Section VI, while Section VIII discusses possible extensions to higher order FIR systems or higher order powers.

II. STATEMENT OF THE PROBLEM

We shall consider a Wiener system composed of a Finite Impulse Response (FIR) system of order m followed by a static nonlinearity of the form x^n: see Figure 1. The system can thus be represented as

$$w(t) = \sum_{k=1}^{m} b_k u(t - k + 1) \qquad (1)$$

$$y_0(t) = w^n(t) \qquad (2)$$

$$y(t) = y_0(t) + v(t) \qquad (3)$$

where $v(t)$ is zero-mean white Gaussian noise with standard deviation σ_v and where $u(t)$ will be assumed to be a zero mean strictly stationary signal with standard deviation σ_u. The system can be rewritten as

$$y(t) = \left[\sum_{k=1}^{m} b_k u(t - k + 1) \right]^n + v(t). \qquad (4)$$

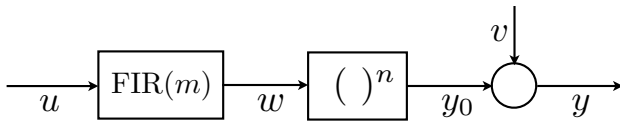The purpose of this paper is to examine the role of the

Fig. 1. Block diagram of the Wiener system

power n and of the coefficients b_k of the FIR model under a variety of assumptions, and to compute optimal inputs in view of the estimation of these parameters b_k. To keep things simple we shall adopt the most classical optimal input design criteria used in identification, namely scalar measures of the asymptotic covariance matrix.

III. AN EXPANDING NONLINEARITY HELPS

In this section we analyze the information matrix for the system of Figure 1 and exhibit the role of the power n of the nonlinear block in the expression of this information matrix, and hence in the achievable precision of the parameter estimates. We first derive the expression of the

information matrix without restricting the linear dynamic block to be an FIR model: thus we momentarily assume that $w(t) = G(z, \theta)u(t)$ where $G(z, \theta)$ is a rational stable transfer function function and the vector θ contains the parameters of the numerator and denominator polynomials of $G(z, \theta)$.

Since v is white noise, the one step ahead predictor of $y(t)$ is $\hat{y}(t, \theta) = G(z, \theta)u(t)$ and its gradient is then

$$\frac{\partial \hat{y}(t, \theta)}{\partial \theta} = nw^{n-1}(t) \frac{\partial G(z, \theta)}{\partial \theta} u(t). \qquad (5)$$

Therefore the average per sample information matrix takes the form

$$I(\theta) = \frac{1}{\sigma_v^2} E \left[\left(\frac{\partial \hat{y}(t, \theta)}{\partial \theta} \right) \left(\frac{\partial \hat{y}(t, \theta)}{\partial \theta} \right)^T \right] \qquad (6)$$

$$= \frac{1}{\sigma_v^2} E \left[\left(\frac{\partial G(z, \theta)}{\partial \theta} u(t) \right) \left(\frac{\partial G(z, \theta)}{\partial \theta} u(t) \right)^T w^{2n-2}(t) \right]$$

where the expectation is with respect to the probability density function of u. We already observe the effect of the power nonlinearity: roughly speaking, the variance of the output $w(t)$ of the linear dynamic system, $E[w^2(t)]$, amplifies the effect of the input signal on each element of the information matrix with a power $n - 1$. When the linear system is a Finite Impulse Response of order m, the per sample information matrix becomes:

$$I_{ij}(\theta) = \frac{n^2}{\sigma_v^2} E \left[u(t - i + 1)u(t - j + 1)w^{2n-2}(t) \right]$$

$$= \frac{n^2}{\sigma_v^2} E \left[\frac{u(t - i + 1)u(t - j + 1)}{(\sum_{k=1}^{m} b_k u(t - k + 1))^2} w^{2n}(t) \right] \qquad (7)$$

Expression (7) shows the effect of the power n of the nonlinearity on the information matrix: each element is multiplied by n^2 times the power of the signal $w(t)$ raised to the power $n - 1$. Thus, an expanding nonlinearity ($n > 1$) greatly increases the information compared to a purely linear system, while a contracting nonlinearity ($n < 1$) decreases the information.

Suppose now that the variance of the noise-free output signal is kept constant when the power n is varied, so that the signal to noise ratio remains fixed:

$$E[w^{2n}(t)] = E[\sum_{k=1}^{m} b_k u(t - k + 1)]^{2n} = \alpha \qquad (8)$$

for some fixed scalar $\alpha > 0$. In view of (7) we expect that, even in such case of constant output variance, the elements of the information matrix will be approximately proportional to the square of the power n, and that the information per parameter decreases when the amplitude of these parameters increases, since the parameter vector appears with a square in the denominator. This is an expected result of the normalization: given that the average per sample covariance matrix is the inverse of the information matrix, $P_\theta = I^{-1}(\theta)$, it implies that the relative uncertainty as measured by $P_\theta/|\theta|^2$ is independent of the norm of the parameter.
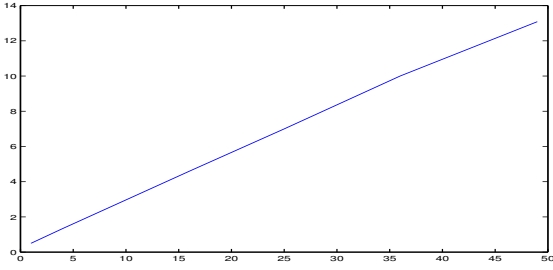
Fig. 2. $I(1,1)$ as a function of n^2

To confirm this we have performed the following simulations. We have taken the FIR(2) model

$$y(t) = [u(t) + u(t-1)]^n + v(t) = w^n(t) + v(t), \quad (9)$$

for powers n going from 1 to 7, where the input signal $u(t)$ was generated as a zero mean Gaussian white noise, scaled as shown in (8), with $\alpha = 1$ for each n. Given the slow convergence of the higher order moments of u, 10^7 realizations of u have been generated to compute the experimental estimates of the elements of the information matrix (7) for each n. Figure 2 shows the $(1,1)$ element of the experimental information matrices as a function of n^2 that resulted from this simulation. It clearly confirms that the elements of $I(\theta)$ are proportional to n^2.

IV. CRAMER-RAO BOUND FOR A STATIC GAIN

To get some more insight into the role of the power n of the nonlinearity and of the parameters of the FIR system, we now specialize these expressions to the special case of a FIR(1) system:

$$y(t) = [\theta u(t)]^n + v(t) = w^n(t) + v(t). \quad (10)$$

Expression (7) becomes

$$I(\theta) = \frac{n^2 \theta^{2n-2}}{\sigma_v^2} E\left[u^{2n}(t)\right]. \quad (11)$$

The Cramér Rao bound for the estimation of θ, on the basis of N data, is

$$P_N(\theta) \approx \frac{\sigma_v^2}{n^2 \theta^{2n-2} N E\left[u^{2n}(t)\right]}. \quad (12)$$

We observe that the achievable precision is inversely proportional to the amplitude of the parameter θ raised to the power $2n-2$, in addition to being inversely proportional to n^2. This is of course because the nonlinear filter amplifies (for $n > 1$) the power of the signal at the output. To make a fair comparison with the linear case, we consider again that the output variance is kept constant at a value α when n is varied. Substituting $E[\theta u(t)]^{2n} = \alpha$ in (11) now yields:

$$I(\theta) = \frac{\alpha n^2}{\sigma_v^2 \theta^2}, \qquad P_N(\theta) \approx \frac{\sigma_v^2 \theta^2}{\alpha n^2 N}. \quad (13)$$

Thus the relative variance of an efficient estimator $\hat{\theta}_N$ will tend for large N to:

$$\frac{Var(\hat{\theta}_N)}{\theta^2} \approx \frac{1}{n^2 N} \frac{\sigma_v^2}{\alpha} \quad (14)$$

where $\frac{\sigma_v^2}{\alpha}$ is the inverse of the signal to noise ratio. This expression is quite enlightening: it shows that, in the case of a constraint on the output variance, the achievable relative precision in the parameter estimate is proportional to the signal to noise ratio and to the square of the power of the static nonlinearity. It shows exactly how an expanding nonlinearity helps in improving the precision of the parameter estimate.

For a contracting nonlinearity

$$y(t) = [\theta u(t)]^{\frac{1}{p}} + v(t), \quad p > 1, \quad (15)$$

the following expressions are obtained:

$$I(\theta) = \frac{\alpha}{p^2 \sigma_v^2 \theta^2}, \qquad \frac{Var(\hat{\theta}_N)}{\theta^2} \approx \frac{p^2}{N} \frac{\sigma_v^2}{\alpha} \quad (16)$$

Hence a contracting nonlinearity increases the uncertainty in the parameter estimate proportionately to the square of the power of the contraction.

We perform Monte-Carlo simulations to validate the expressions (13) by taking a range of values for θ and n. The white noise v has unit variance, while the variance of $u(t)$ is normalized so that the expected variance of the noiseless output signal is equal to 1, i.e. $E[\theta u(t)]^n = 1$ for all n. In each simulation, we first generate a zero mean unit variance Gaussian signal $\epsilon(t)$ and we subsequently divide $\epsilon(t)$ by the $2n$-th root of the theoretical variance of $y_0(t) = [\theta \epsilon(t)]^n$. For a Gaussian signal, we have

$$E[y_0(t)]^2 = \theta^{2n} E[\epsilon^{2n}(t)] = \theta^{2n} \sigma_\epsilon^{2n} (2n-1)!! \quad (17)$$

where the double factorial $(2n-1)!!$ is defined as $(2n-1)!! = (2n-1)(2n-3) \ldots 1$. Thus, the input signal $u(t)$ is obtained from the unit variance white noise $\epsilon(t)$ as follows:

$$u(t) = \frac{\epsilon(t)}{|\theta| \sqrt[2n]{(2n-1)!!}} \quad (18)$$

The parameter estimate is obtained by first estimating β by Least Squares regression from the equations

$$y(t) = \beta x(t) + v(t) \quad (19)$$

with $x(t) = u^n(t)$, and then taking $\hat{\theta}_N = \sqrt[n]{\hat{\beta}_N}$.

Figure 3 presents the values of the experimental information $I(\theta)$ obtained by Monte-Carlo simulations for 3 different values of the parameter θ and for powers n ranging from 1 to 9, compared with the theoretical values obtained from (13). For the sake of clarity, the information is presented in a logarithmic scale, i.e. $db(I(\theta))$. It can be shown that the convergence of the experimental moments $\frac{1}{N} \sum_{k=1}^{N} u^n(k)$ to their theoretical values $E[u^n(t)]$ becomes extremely slow when n increases; as a consequence, these simulations were produced with large values of $N = 10,000$.
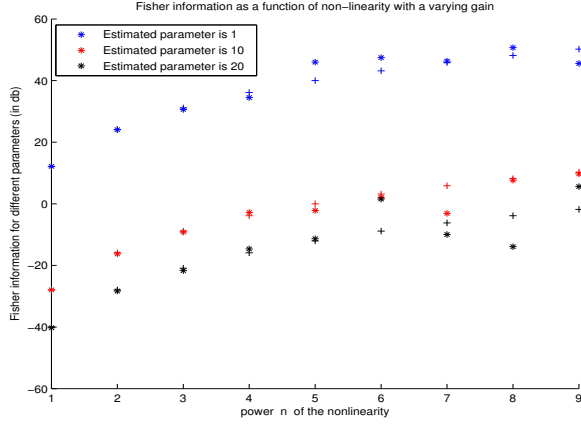
Fig. 3. Fisher information (in db) as a function of the nonlinearity and of the parameter; theoretical (+) and simulated values (∗).

V. OPTIMAL INPUT DESIGN FOR AN FIR(2) SYSTEM

Having established some interesting properties of the information matrix for the class of Wiener systems considered in Figure 1, we now address the optimal input design problem for the same class of systems. To simplify matters, we first analyze the case of a FIR(2) system followed by a static nonlinearity that takes the square of its input; extensions to FIR systems of higher order and to higher powers of the static nonlinear block will be discussed in Section VIII. Thus, the system can be represented as

$$y(t) = [b_1 u(t) + b_2 u(t-1)]^2 + v(t). \qquad (20)$$

Defining $\theta \triangleq [b_1 \ b_2]^T$ yields

$$\frac{\partial \hat{y}(t, \theta)}{\partial \theta} = \begin{pmatrix} 2[b_1 u(t) + b_2 u(t-1)]u(t) \\ 2[b_1 u(t) + b_2 u(t-1)]u(t-1) \end{pmatrix} \qquad (21)$$

Therefore the elements $I_{ij}(\theta)$ of the average per sample information matrix can be written as

$$
\begin{aligned}
I_{11} &= \frac{4}{\sigma_v^2}(b_1^2 m_u^{40} + 2b_1 b_2 m_u^{31} + b_2^2 m_u^{22}) \\
I_{12} &= I_{21} = \frac{4}{\sigma_v^2}(b_1^2 m_u^{31} + 2b_1 b_2 m_u^{22} + b_2^2 m_u^{13}) \\
I_{22} &= \frac{4}{\sigma_v^2}(b_1^2 m_u^{22} + 2b_1 b_2 m_u^{13} + b_2^2 m_u^{04})
\end{aligned} \qquad (22)
$$

where we have introduced the following notation for the moments of $u(t)$:

$$m_u^{ij} \triangleq E[u^i(t)u^j(t-1)]. \qquad (23)$$

In the linear case ($n = 1$) the information matrix for an FIR(2) system takes the form

$$I(\theta) = \frac{1}{\sigma_v^2} \begin{bmatrix} E[u^2(t)] & E[u(t)u(t-1)] \\ E[u(t)u(t-1)] & E[u^2(t)] \end{bmatrix} \qquad (24)$$

which depends only on the autocovariance function of the input at lag zero and one. The optimal input is then entirely set by the power spectrum of the input signal as is well

known from the literature. Whatever the optimal design criterion, the optimization can then be performed with respect to $E[u^2(t)]$ and $E[u(t)u(t-1)]$ subject to the constraint that $I(\theta) \geq 0$ in (24), which guarantees that $E[u^2(t)]$ and $E[u(t)u(t-1)]$ are the first two autocovariance elements of a valid spectrum Φ_u. The optimal input design problem is then solved in two steps. First the optimal criterion is minimized with respect to $E[u^2(t)]$ and $E[u(t)u(t-1)]$ subject to $I(\theta) \geq 0$ and other possible constraints imposed by the designer; a large number of such problems can be formulated as convex optimization problems subject to Linear Matrix Inequality (LMI) constraints. The next step consists in generating an input sequence whose spectrum matches these first two autocovariance elements.

In the nonlinear Wiener system of this paper, we observe that the information matrix depends on higher order moments of the input signal. In the particular case studied in this section, it depends on four parameters: $m_u^{40}, m_u^{31}, m_u^{22}, m_u^{13}$, observing that by the stationarity assumption we have $m_u^{40} = m_u^{04}$. Thus, as observed in [2], it might be tempting to solve the optimal input design problem with respect to these four parameters. However, there are two constraints:
(i) they have to satisfy the constraint $I(\theta) \geq 0$ where $I(\theta)$ depends on $m_u^{40}, m_u^{31}, m_u^{22}, m_u^{13}$ through (22);
(ii) there must exist a valid probability distribution $p(u(t), u(t-1))$ that generates these moments.

Problem (i) is not too difficult to address since $I(\theta) \geq 0$ is an LMI in these unknown parameters, even if it depends on the parameters of the unknown FIR system. This is the traditional paradigm of optimal experiment design: the optimal solution always depends on the unknown system. As for problem (ii), to the best ot the authors' knowledge, this is still an open problem.

We observe that, given the assumption on strict stationarity of the input $u(t)$ that we have made throughout the paper, the optimal solution of an input design problem is completely characterized by the knowledge of the joint probability density function $p(u(t), u(t-1))$.

A suboptimal procedure for the design of optimal inputs for nonlinear systems, suggested in [2], would be to parametrize the joint density $p(u(t), u(t-1))$ as

$$p(u(t), u(t-1)) = \sum_{k=1}^{l} \gamma_k p_k(u(t), u(t-1)) \qquad (25)$$

where the $p_k(u(t), u(t-1))$ are preselected distributions and the parameters γ_k are free variables subject to $\sum_{k=1}^{l} \gamma_k = 1$. In such case, the moments m_u^{ij} appearing in the information matrix (see (22) can be written as

$$m_u^{ij} = \sum_{k=1}^{l} \gamma_k \iint u^i(t) u^j(t-1) p_k(u(t), u(t-1)) du(t) du(t-1) \qquad (26)$$

The constraint $I(\theta) \geq 0$ now becomes an LMI in the parameters $\gamma_k, k = 1, \ldots, l$.

In this paper we examine another suboptimal solution in which the class of inputs is restricted to be Gaussian.

Even though our analysis applies to different optimal design criteria, for pedagogical reasons we consider that the goal is to design an input that maximizes the determinant of the information matrix $I(\theta)$.

VI. THE CASE OF GAUSSIAN INPUTS

Consider thus that we restrict the input to be a strictly stationary zero-mean Gaussian signal $u(t)$ with variance σ_u^2 and covariance function $E[u(t)u(t - \tau)] = \rho_u(\tau)\sigma_u^2$. Denoting $\rho \triangleq \rho_u(1)$, the moments m_u^{ij} tcan be written:

$$
\begin{aligned}
m_u^{40} &= 3\sigma_u^4 \\
m_u^{31} &= m_u^{13} = 3\rho\sigma_u^4 \\
m_u^{22} &= (1 + 2\rho^2)\sigma_u^4.
\end{aligned}
$$

The information matrix can then be rewritten as

$$
I_{11} = \frac{4\sigma_u^4}{\sigma_v^2}[3b_1^2 + 6b_1b_2\rho + b_2^2(1 + 2\rho^2)] \tag{27}
$$

$$
I_{12} = I_{21} = \frac{4\sigma_u^4}{\sigma_v^2}[3b_1^2\rho + 2b_1b_2(1 + 2\rho^2) + 3b_2^2\rho]
$$

$$
I_{22} = \frac{4\sigma_u^4}{\sigma_v^2}[b_1^2(1 + 2\rho^2) + 6b_1b_2\rho + 3b_2^2]
$$

The determinant of $I(\theta)$ then takes the following form:

$$
\begin{aligned}
detI(\theta) = 48\frac{\sigma_u^8}{\sigma_v^4}\Big[&-4b_1^2b_2^2\rho^4 - 4b_1b_2(b_1^2 + b_2^2)\rho^3 \\
&-(b_1^2 - b_2^2)^2\rho^2 + 4b_1b_2(b_1^2 + b_2^2)\rho + (b_1^2 + b_2^2)^2\Big]
\end{aligned} \tag{28}
$$

The only design parameter for the maximization of the determinant of $I(\theta)$ is the autocorrelation coefficient ρ, with the constraint that $|\rho| < 1$. In order to find a maximizing value, we set the derivative of $detI(\theta)$ with respect to ρ to zero:

$$
\begin{aligned}
\frac{\partial detI(\theta)}{\partial \rho} = 48\frac{\sigma_u^8}{\sigma_v^4}\Big[&-16b_1^2b_2^2\rho^3 - 12b_1b_2(b_1^2 + b_2^2)\rho^2 \\
&-2(b_1^2 - b_2^2)^2\rho + 4b_1b_2(b_1^2 + b_2^2)\Big] = 0.
\end{aligned} \tag{29}
$$

The maximum of $detI(\theta)$ is obtained for those roots of (29), if any, that are in $(-1, 1)$ and for which $\frac{\partial^2 detI(\theta)}{\partial \rho^2} < 0$. We have

$$
\begin{aligned}
\frac{\partial^2 detI(\theta)}{\partial \rho^2} = 96\frac{\sigma_u^8}{\sigma_v^4}\Big[&-24b_1^2b_2^2\rho^2 - 12b_1b_2(b_1^2 + b_2^2)\rho \\
&-(b_1^2 - b_2^2)^2\Big].
\end{aligned} \tag{30}
$$

An optimal solution can always be generated as

$$
u(t) = \rho u(t - 1) + \epsilon(t) \tag{31}
$$

where $\epsilon(t)$ is white Gaussian noise with mean zero and variance σ_ϵ^2. The variance of the input signal is then $\sigma_u^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$.

VII. SIMULATIONS WITH GAUSSIAN INPUTS

To illustrate the optimal input design for the Wiener process of Figure 1 we have taken the following system:

$$
y(t) = [5u(t) + 2u(t - 1)]^2 + v(t), \tag{32}
$$

with $\sigma_v^2 = 0.25$. The input is generated by (31) with $\epsilon(t)$ being also zero mean white Gaussian noise whose variance is taken such that the input signal $u(t)$ has variance $\sigma_u^2 = 1$.

As stated above, the only design parameter for the maximization of the determinant of the information matrix is the autocorrelation ρ. The maximization of the determinant of the intormation matrix $I(\theta)$ given by (28) with respect to ρ subject to the constraint $\sigma_u^2 = 1$, yields $\rho = 0.4321$.

In order to evaluate the behaviour of the determinant of $I(\theta)$ as a function of ρ, we have plotted the value of $detI(\theta)$ given by (28) with $b_1 = 5$ and $b_2 = 2$ as a function of ρ. We have also computed the determinant of $I(\theta)$ by simulations, using the expression (22), in which the m_u^{ij} have been replaced by their experimental expressions $\hat{m}_u^{ij} = \frac{1}{N}\sum_{t=1}^{N} u^i(t)u^j(t - 1)$.

Figure 4 shows the determinant of the information matrix as a function of the autocorrelation ρ: the black curve (+) is based on the theoretical expression (28), the red curve (∗) is based on the simulated information matrix computed from 50000 data, while the cyan curve (-) is computed in the same way as the red curve but with only 5000 data. The convergence of the expressions $\frac{1}{N}\sum_{t=1}^{N} u^4(t)$, $\frac{1}{N}\sum_{t=1}^{N} u^3(t)u(t - 1)$ and $\frac{1}{N}\sum_{t=1}^{N} u^2(t)u^2(t - 1)$ to their theoretical values $3\sigma_u^4$, $3\rho\sigma_u^4$ and $(1 + 2\rho^2)\sigma_u^4$, respectively, is very slow, which explains the difference between the red and cyan curves in Figure 4.
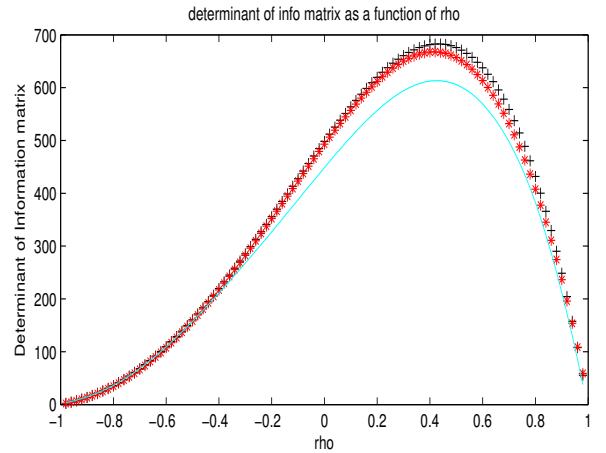

Fig. 4. Determinant of the information matrix as a function of ρ. Black(+): theoretical information matrix, red(∗): simulated with 50,000 data; cyan (-): simulated with 5,000 data.

In order to validate these findings, we have performed another Monte Carlo simulation in which we have estimated the parameters b_1 and b_2 by nonlinear least squares (using the Matlab function fminsearch) and computed their experimental covariance for inputs generated by (31) and with ρ ranging in the interval $(-1, 1)$. The estimation of the covariance was

based on 2000 Monte Carlo simulations, each using 2000 data. In order to be consistent with the results of Section VI, the variance of ϵ in (31) was adjusted so that the variance of u was independent of ρ. Figure 5 shows the determinant of the experimental covariance as a function of ρ, which is consistent with the results of Figure 4.
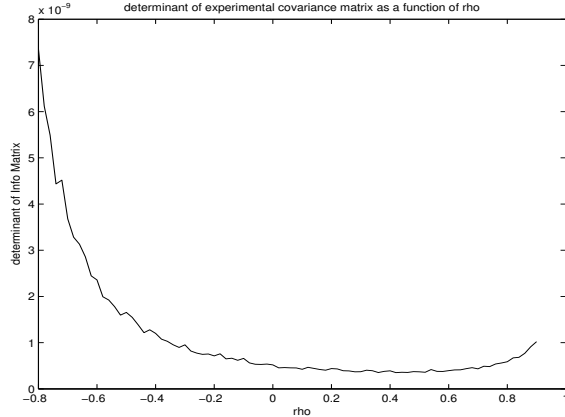

Fig. 5. Determinant of the experimental covariance matrix as a function of ρ. Black(+): theoretical information matrix, red(∗): simulated with 50,000 data; cyan (-): simulated with 5,000 data.

VIII. EXTENSIONS

In this section we examine the extension of our results for FIR(2) systems followed by a square nonlinearity to the cases of FIR(m) systems and higher powers of the nonlinear block. We also consider alternative families of input signals.

We first observe that there are no theoretical difficulties in extending the expressions for the information matrix from (22) to corresponding expressions with more than two parameters and higher powers of n than 2. However, we note that the number of design parameters increases quickly with m and n. For $m = 2$ and increasing powers of n, the solution will only depend on higher order moments of the form m_u^{ij} as defined in (23) involving powers of $u(t)$ and $u(t-1)$ only, with $i, j = 0, \ldots, n$ and $i + j = n$. This is still a reasonably tractable problem. However, the complexity of the problem increases rapidly for longer FIR models (i.e. $m > 2$), since all moments of the form $E[u^i(t-k+1)u^j(t-l+1)]$ must be considered for $i, j = 0, \ldots, n$ with $i + j = n$, and for $k, l = 1, \ldots, m$.

If the inputs are restricted to be Gaussian, these problems are again very tractable, since these higher order moments are simple functions of σ_u and the autocovariance function of u up to lag m.

From the previous study it turned out that the optimal input design depends on the higher order moments of the joint probability density function. A full optimal design should optimize over all these parameters, but as explained before we cannot handle the full problem at this moment. A simplified problem is to optimize both the power spectrum and the amplitude distribution of the excitation. Even this solution is not feasible yet in practice, because to the best

of the authors' knowledge, there are no algorithms available where the user can set independently the power spectrum and the amplitude distribution of a random sequence (e.g. generate a uniform distribution with a given power spectrum). Generating deterministic excitation signals offers an intermediate solution to the problem. In [6] it is shown that it is possible to generate periodic excitations (called multisines) where the user has full control over the power spectrum. The phases are selected such that the amplitude distribution of the excitation approximates a user defined amplitude distribution. This idea can be reused within the context of solving the generalized optimal input design problem, but more research is needed here to make hard claims.

IX. CONCLUSIONS

This work is a contribution to the emerging field of experiment design for structured nonlinear systems. We have examined the class of Wiener systems made up of a linear FIR system followed by a nonlinear power function. We have shown that nonlinearities can be beneficial to the estimation of parameters in the case of expanding powers. As for the experiment design aspects, we have illustrated, via a simple FIR(2) system followed by a static square power function, the kind of calculations that are required for the computation of optimal inputs. Almost surely, one will only be able to compute suboptimal solutions by restricting the class of considered inputs. We have shown that, for Gaussian inputs, reasonably simple expressions are obtained leading to simple suboptimal solutions.

REFERENCES

[1] M. Gevers, X. Bombois, R. Hildebrand, and G. Solari. Optimal experiment design for open and closed-loop system identification. *Communications in Information and Systems*, 11(3):197–224, 2011.

[2] H. Hjalmarsson, J. Martensson, and B. Ninness. Optimal input design for identification of non-linear systems: Learning from the linear case. In *American Control Conference*, pages 1572–1576, July 2007.

[3] C.A. Larsson, H. Hjalmarsson, and C.R. Rojas. On optimal input design for nonlinear fir systems. In *IEEE Conference on Decision and Control*, pages 7220–7225, Atlanta, GA, USA, December 2010.

[4] Lu Lu and Bin Yao. Experimental design for identification of nonlinear systems with bounded uncertainties. In *Proc. American Control Conference*, pages 4504–4509, Baltimore, Maryland, USA, June 30 - July 2 2010.

[5] R. K. Mehra. Optimal input signals for parameter estimation in dynamic systems - survey and new results. *IEEE Trans. on Automatic Control*, AC-19(6):753–768, December 1974.

[6] J. Schoukens and T. Dobrowiecki. Design of broadband excitation signals with a user imposed power spectrum and amplitude distribution. In IEEE, editor, *Proc. IEEE Conf. on Instrumentation and Measurement Technology*, pages 1002–1005, 1998.

[7] T.L. Vincent, C. Novara, K. Hsu, and K. Poolla. Input design for structured nonlinear system identification. In *15th IFAC Symposium on System Identification*, pages 174–179, Saint Malo, France, July 2009.