

Nonlinear Structure Identification With Application to Wiener-Hammerstein Systems

D.J. Leith¹, W.E. Leithead^{1,2}, R. Murray-Smith^{1,3}

¹Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland

²Dept. of Electronics & Electrical Engineering, University of Strathclyde, Glasgow G1 1QE, U.K.

³Dept. of Computing Science, University of Glasgow, Glasgow G12 8QQ, U.K.

Abstract

While there exists a substantial literature on the identification of Hammerstein and Wiener models, the identification of Wiener-Hammerstein models has received considerably less attention yet this is a model class of very great practical importance. This paper proposes an elegant approach to estimating Wiener-Hammerstein systems from measured data.

1. Introduction

While there exists a substantial literature on the identification of Hammerstein and Wiener models, the identification of Wiener-Hammerstein models has received considerably less attention (recent papers include Bai 1998, 2002; see also the survey of Haber & Unbehauen 1990) yet this is a model class of very great practical importance. In this paper we consider the identification of transversal Wiener-Hammerstein systems

$$\begin{aligned} x_i &= a_n r_{i-n} + \dots + a_o r_i \\ z_i &= f(x_i) \\ y_i &= b_m z_{i-m} + \dots + b_o z_i \end{aligned} \quad (1)$$

Assume that $N > n+m$ noisy input-output measurements $\{(r_i, \hat{y}_i)\}_{i=1}^N$ are available, where $\hat{y}_i = y_i + n_i$, n_i zero mean Gaussian noise. One common strategy is to seek to directly fit the parameters in (1) to the measured data. An alternative strategy, however, is to adopt an indirect approach as follows. Let \mathbf{R}_i denote the delayed input vector

$$\mathbf{R}_i = \begin{bmatrix} r_{i-(n-m)} \\ r_{i-(n-m-1)} \\ \vdots \\ r_i \end{bmatrix} \quad (2)$$

and $D = \{\mathbf{R}_i, i=n+m, \dots, N\}$ be the set of measured delayed input vectors. Reformulating the dynamics in terms of \mathbf{R}_i and y_i yields

$$y_i = F(\mathbf{R}_i) \quad (3)$$

where

$$\begin{aligned} F(\mathbf{R}_i) &= b_m f(\mathbf{M}_{m+1} \mathbf{R}_i) + \dots + b_o f(\mathbf{M}_1 \mathbf{R}_i) \\ \mathbf{M} &= \begin{bmatrix} a_n & \dots & a_o & 0 & \dots & 0 \\ 0 & a_n & \dots & a_o & 0 & \dots & 0 \\ & & & & \ddots & & \\ 0 & \dots & 0 & a_n & \dots & a_o \end{bmatrix} \end{aligned} \quad (4)$$

and \mathbf{M}_k , $k=1..m+1$ denotes the k^{th} row of \mathbf{M} . Without loss it is assumed that $f(0)=0$, $\nabla f(0)=1$. We also assume that $\nabla^2 f(\mathbf{R}_i)$

is *not* identically zero¹ $\forall \mathbf{R}_i \in D$, $a_n=1$ and that at least one of the coefficients b_k , $k=0..m$ is non-zero. Since the input measurements are noise free, inference of the map F in (3) relating output y_i to input \mathbf{R}_i may be formulated as a regression problem. Many methods exist which might be employed to estimate F . In this paper, we employ a non-parametric Gaussian process prior approach; see Appendix. Our objective is then to determine a minimal realisation (1) of the system from this non-minimal input-output formulation. The derivation of minimal realisations of nonlinear systems is well known to be a difficult task in general. However, the minimal realisation task for Wiener-Hammerstein systems is potentially more tractable. Indeed, the key issue is to determine the matrix \mathbf{M} and it is this which is the focus of the present paper. This matrix effectively partitions the linear dynamics into the input filter and the output filter elements. Once \mathbf{M} is known, the identification task reduces to the estimation of the Hammerstein model relating the output y_i to the transformed input $\mathbf{M}\mathbf{R}_i$. This is a relatively straightforward problem for which standard techniques exist.

Notation: The notation used is standard. For a matrix $\mathbf{M} \in \mathfrak{R}^p$, $\text{null}(\mathbf{M})$ denotes the null space of \mathbf{M} , *i.e.* $\text{null}(\mathbf{M}) = \{v \in \mathfrak{R}^p: \mathbf{M}v=0\}$, and $\text{range}(\mathbf{M})$ denotes the range space of \mathbf{M} , *i.e.* $\text{range}(\mathbf{M}) = \{v \in \mathfrak{R}^p: \mathbf{M}v \neq 0\}$. We let $\text{basis}\{V\}$ denote any basis of the subspace $V \subseteq \mathfrak{R}^p$ *i.e.* any orthonormal set of vectors spanning V . For a twice differentiable map $\mathbf{F}: \Delta \subseteq \mathfrak{R}^p \rightarrow \mathcal{P} \subseteq \mathfrak{R}^l$, $\mathbf{H}_{\mathbf{F}}(\mathbf{z})$ denotes the Hessian $\mathbf{H}_{\mathbf{F}}(\mathbf{z}) = [\nabla(\nabla F_1(\mathbf{z}))^T \dots (\nabla F_q(\mathbf{z}))^T]^T$ with F_i denoting the i^{th} element of the map \mathbf{F} .

2. Deterministic Case

Assume, initially, that the nonlinear map F in (3) is known. We assume also that the associated Hessian map, $\mathbf{H}_{\mathbf{F}}$, is

¹ This is a form of identifiability condition: if $\nabla^2 f(\mathbf{R}_i)=0 \forall \mathbf{R}_i \in D$, then on the basis of the measured data it is readily verified that f cannot be distinguished from a linear function.

known. The latter assumption is little more than a smoothness assumption on F as the Hessian may be immediately derived provided F is specified in a closed-form manner. For example, in a classical regression context where F might be expressed as a sum of basis functions,

$$F(\mathbf{R}) = \sum_{k=1}^{n_\phi} \theta_k \phi_k(\mathbf{R}) \quad (5)$$

we have that

$$\mathbf{H}_F(\mathbf{R}) = \sum_{k=1}^{n_\phi} \theta_k \nabla(\nabla\phi_k(\mathbf{R}))^T \quad (6)$$

Here, the basis functions ϕ_k might for example be Gaussian, sigmoidal, polynomial *etc.* Notice that for a given model structure, once the parameters θ_k of the map F are specified the Hessian \mathbf{H}_F is also completely specified.

With regard to the Wiener-Hammerstein system (3), it is readily verified that

$$\mathbf{H}_F(\mathbf{R}) = \mathbf{M}^T \nabla^2 f(\mathbf{R}) \begin{bmatrix} \mathbf{b}_m \mathbf{M} \\ \vdots \\ \mathbf{b}_o \mathbf{M} \end{bmatrix} \quad (7)$$

where $\mathbf{R} \in \mathfrak{R}^{n+m+1}$. Hence, provided $\nabla^2 f(\mathbf{R}) \neq 0$ it follows immediately that

$$\mathbf{H}_F(\mathbf{R})\mathbf{v} = 0 \Leftrightarrow \mathbf{M}\mathbf{v} = 0 \quad \forall \mathbf{v} \in \mathfrak{R}^{n+m} \quad (8)$$

i.e. $\text{null}(\mathbf{H}_F(\mathbf{R})) = \text{null}(\mathbf{M})$. Since $\nabla^2 f(\mathbf{R})$ is not everywhere zero in D (or else the system is trivially linear), this property may be expressed as

$$\bigcap_{\mathbf{R} \in D} \text{null}(\mathbf{H}_F(\mathbf{R})) = \text{null}(\mathbf{M}) \quad (9)$$

This observation is key importance in the sequel as it implies that the matrix \mathbf{M} may be determined by direct inspection of the Hessian of the input-output map F .

Remarks

(i) The matrix \mathbf{M} has a specific block diagonal structure. By inspection, the coefficients, \mathbf{a}_k , of the input filter and the delay taps of the output filter can be directly inferred via (4). Identification of the remaining system elements is now relatively straightforward. For example, once the input filter is known, the output filter can be inferred from the transfer function of the linearised dynamics about any equilibrium point. With the input and output linear dynamics known (*i.e.* the coefficients \mathbf{a}_k and \mathbf{b}_k) the scalar nonlinear function f might be determined directly from the curve

$$\bar{y} = \left(\sum_{k=0}^m \mathbf{b}_k \right) f(\bar{x}) \quad (10)$$

relating the equilibrium output \bar{y} to $\bar{x} = \sum_{k=0}^n \mathbf{a}_k \bar{r}$, where \bar{r} is the associated equilibrium input value.

(ii) In general, the matrix \mathbf{M} is not uniquely determined by the input-output properties of the system: we cannot distinguish between structured matrices \mathbf{M} with identical null spaces. This arises because scalings in the coefficients \mathbf{a}_k can be absorbed in the function f . This degree of freedom is removed here through the constraint that \mathbf{a}_n is unity.

3. Nonlinear System Identification

The structural analysis in section 2 is deterministic. In this section we consider the extension to a probabilistic context. We assume that probabilistic (stochastic process) descriptions are available of $F(\mathbf{R})$ and its directional derivative $\mathbf{H}_F(\mathbf{R})\mathbf{v}$ (for any given direction \mathbf{v}). The stochastic process description of the directional derivative $\mathbf{H}_F(\mathbf{R})\mathbf{v}$ can be determined from the stochastic process description of $F(\mathbf{R})$ (*e.g.* in the case of a Gaussian process, the mean and covariance of $\mathbf{H}_F(\mathbf{R})\mathbf{v}$ are appropriate derivatives of the mean and covariance of $F(\mathbf{R})$ – see Appendix). It is perhaps worth emphasising that this certainly does not require differentiation of the raw, noisy data. The latter is, of course, highly inadvisable.

Remark It is important to note that, while most previous work on the identification of Wiener-Hammerstein systems has assumed that the nonlinear function f has a specific parametric structure (typically polynomial), stochastic process descriptions do not necessarily require the imposition of a parametric structure. Non-parametric descriptions (*e.g.* Green & Silverman 1994, Neal 1996, Williams 1998) are characterised by drawing inferences directly from the measured data using smoothness information but *without* assuming an underlying parameterisation. (Various forms of smoothness assumption are typically employed: any specific assumption may of course be more or less appropriate in a particular application context). An example of a non-parametric non-linear description is a Gaussian process prior model used here: see Appendix.

3.1 Estimating \mathbf{M}

The condition $\bigcap_{\mathbf{R} \in D} \text{null}(\mathbf{H}_F(\mathbf{R})) = \text{null}(\mathbf{M})$ is equivalent to

- (i) $\mathbf{H}_F(\mathbf{R})\mathbf{v} = 0 \quad \forall \mathbf{v} \in \text{basis}\{\Psi_1\}, \mathbf{R} \in D$
- (ii) $\mathbf{H}_F(\mathbf{R})\mathbf{v} \neq 0 \quad \forall \mathbf{v} \in \text{basis}\{\Psi_{nl}\}, \mathbf{R} \in D$

where Ψ_1 denotes $\text{null}(\mathbf{M})$ and Ψ_{nl} denotes $\text{range}(\mathbf{M})$. Assuming that the dimension of Ψ_1 is known and that $\mathbf{H}_F(\mathbf{R})\mathbf{v}$ is now described by a probability distribution³, the probability distribution corresponding to (i) is

$$p_M = p(\mathbf{H}_F(\mathbf{R})\mathbf{v} = 0; \mathbf{v} \in \text{basis}\{\Psi_1\}, \mathbf{R} \in D) \quad (11)$$

Exploiting the block diagonal structure of \mathbf{M} for Wiener-Hammerstein systems, we have that $\text{basis}\{\Psi_1\}$ may be expressed as

$$\text{basis}\{\Psi_1\} = \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{m+1}\} \quad (12)$$

where

² Note that, with an abuse of notation, we use $\mathbf{H}_F(\mathbf{R})$ to also refer to the stochastic process but the meaning should be clear.

³ While $\mathbf{H}_F(\mathbf{R})\mathbf{v}$ is a matrix, the elements are stacked into a vector for working with standard multivariate probability descriptions.

$$\begin{aligned}
\mathbf{M}_1 &= [a_n \ \cdots \ a_o \ 0 \ \cdots \ 0]^T \\
\mathbf{M}_2 &= [0 \ a_n \ \cdots \ a_o \ 0 \ \cdots \ 0]^T \\
&\vdots \\
\mathbf{M}_{m+1} &= [0 \ \cdots \ 0 \ a_n \ \cdots \ a_o \ 0]^T
\end{aligned} \tag{13}$$

Coefficients a_k , $k=1..n$ for which the probability

$$p_M = \mathbb{P}(\mathbf{H}_F(\mathbf{R})\mathbf{v} = 0; \mathbf{v} \in \{\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_{m+1}\}, \mathbf{R} \in \mathcal{D}) \tag{14}$$

is maximal therefore provide an estimate of \mathbf{M}

The foregoing assumes that the dimension of Ψ_1 is known *i.e.* the order m of the output filter. In using (13) we also assume knowledge of the order n of the input filter. This may be relaxed to an assumption only on the overall model order $n+m$ (thereby avoiding assumptions as to partitioning of the dynamics between input and output filters). We proceed as follows. We sequentially increase the dimension of Ψ_1 , re-estimating the basis. Let Ψ_1^i denote the estimated sub-space of dimension i and let Ψ_1 and Ψ_{n1} denote the true sub-spaces. We must have that $\Psi_1^i \cap \Psi_{n1} \neq \emptyset$ for $i > m$. Consequently, as i is increased beyond the true dimension m , we can expect the value of p_M to abruptly decrease (since $\mathbf{H}_F(\mathbf{z})\mathbf{v} \neq 0 \ \forall \ \mathbf{v} \in \text{basis}\{\Psi_{n1}\}$) and can thereby infer the dimension from the data.

4. Example

Consider the Wiener-Hammerstein nonlinear system illustrated in Figure 1. Reformulating the dynamics in terms of the measured variables (input r and output y) yields

$$y(t_n) = 0.3(\mathbf{M}_1 \mathbf{R})^2 + 0.165(\mathbf{M}_3 \mathbf{R})^2 \tag{15}$$

where $\mathbf{R} = [r(t_i) \ r(t_{i-1}) \ r(t_{i-2}) \ r(t_{i-3})]^T$ with

$$\mathbf{M} = \begin{bmatrix} 0.9184 & 0.3674 & 0 & 0 \\ 0 & 0.9184 & 0.3674 & 0 \\ 0 & 0 & 0.9184 & 0.3674 \end{bmatrix} \tag{16}$$

The plant output in response to a Gaussian input is measured: data is collected for 15 seconds with a sampling interval of 0.1 seconds (150 data points). Gaussian white noise of standard deviation 0.1 units is added to the output measurement (the underlying signal has a peak magnitude of 0.5, so this represents a substantial level of noise). A non-parametric Gaussian Process prior model is used with explanatory variables/regressors $[r(t_i) \ r(t_{i-1}) \ r(t_{i-2}) \ r(t_{i-3})]^T$ and model output $y(t_i)$. The change in p_M as the dimension of Ψ_1 (the dimension of the output filter m) is varied indicates a dimension of two as expected. The associated estimate of \mathbf{M} is

$$\hat{\mathbf{M}} = \begin{bmatrix} 0.9292 & 0.3694 & 0 & 0 \\ 0 & 0.9292 & 0.3694 & 0 \\ 0 & 0 & 0.9292 & 0.3694 \end{bmatrix} \tag{17}$$

The estimate evidently agrees well with the true \mathbf{M} , particularly in view of the low signal to noise ratio and small number of data points on which it is based (150 points from a four dimensional map).

5. Conclusions

The is paper investigates new ways of estimating Wiener-Hammerstein models from measured data. A constructive algorithm is proposed and its application is illustrated in a simple example.

Acknowledgements

This work was supported by Science Foundation Ireland grant 00/PI.1/C067, by the EC through EC TMR grant HPRNCT-1999-00107, and EPSRC grants GR/M76379/01 and GR/R15863/01.

Appendix – Non-parametric Gaussian process priors

Consider a smooth function $f(\cdot)$ dependent on $\mathbf{z} \in \mathcal{D} \subseteq \mathcal{R}^p$. To avoid cumbersome notation, f is scalar. Suppose N measurements, $\{(\mathbf{z}_i, y_i)\}_{i=1}^N$, of the value of the function with additive Gaussian white measurement noise, *i.e.* $y_i = f(\mathbf{z}_i) + n_i$, are available and denote them by M . It is of interest here to use this data to learn the mapping $f(\mathbf{z})$ or, more precisely, to determine a probabilistic description of $f(\mathbf{z})$. Note that this is a regression formulation and it is assumed the input \mathbf{z} is noise free⁴. The probabilistic description of the function, $f(\mathbf{z})$, adopted is the stochastic process, $f_{\mathbf{z}}$. By necessity, the probability distributions of $f_{\mathbf{z}}$ for every choice of value of $\mathbf{z} \in \mathcal{D}$ are required together with the joint probability distributions of $f_{\mathbf{z}_i}$ for every choice of finite sample, $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$, from \mathcal{D} , for all $k > 1$. Of course, the joint probability distributions of lower dimensionality must be the marginal distributions of those of higher dimensionality. The $E[f_{\mathbf{z}}]$ as \mathbf{z} varies is interpreted as a fit to $f(\mathbf{z})$.

In the Bayesian probability context, the prior belief is placed directly on the probability distributions describing $f_{\mathbf{z}}$ which are then conditioned on the observations, M , to determine the posterior probability distributions. In Gaussian Process prior model it is assumed that the prior probability distributions for the $f_{\mathbf{z}}$ are all Gaussian with zero mean (in the absence of any evidence the value of $f(\mathbf{z})$ is as likely to be positive as negative). To complete the statistical description, requires only a definition of the covariance function $C(f_{\mathbf{z}_i}, f_{\mathbf{z}_j}) = E[f_{\mathbf{z}_i}, f_{\mathbf{z}_j}]$, for all \mathbf{z}_i and \mathbf{z}_j . The resulting posterior probability distributions are also Gaussian. The Gaussian assumption may seem strangely restrictive initially, but recall that this is simply a prior on the relevant stochastic process space and so places few inherent restrictions on the class of nonlinear functions that can be modelled. Indeed, it can be shown that the result is, in fact, a Bayesian form of kernel regression model (Green & Silverman 1994) subsuming RBF, spline and many neural network models (Williams 1998).

⁴No attempt to being made here to propagate a Gaussian or other distribution through a nonlinear function.

This model is used to carry out inference as follows. We have that $p(\mathbf{f}_z | \mathbf{M}) = p(\mathbf{f}_z, \mathbf{M}) / p(\mathbf{M})$ where $p(\mathbf{M})$ acts as a normalising constant. Hence,

$$p(\mathbf{f}_z | \mathbf{M}) \propto \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{f}_z & \mathbf{F}^T \end{bmatrix} \begin{bmatrix} \Lambda_{11} & \mathbf{L}_{21}^T \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}_z \\ \mathbf{F} \end{bmatrix} \right]$$

where $\mathbf{F} = [y_1, \dots, y_N]^T$, Λ_{11} is $C(\mathbf{f}_z, \mathbf{f}_z)$, the ij^{th} element of the covariance matrix \mathbf{L}_{22} is $C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j})$ and the i^{th} element of vector \mathbf{L}_{21} is $C(\mathbf{f}_{z_i}, \mathbf{f}_z)$. Both Λ_{11} and \mathbf{L}_{21} depend on \mathbf{z} . Applying the partitioned matrix inversion lemma, it follows that

$$p(\mathbf{f}_z | \mathbf{M}) \propto \exp \left[-\frac{1}{2} (\mathbf{f}_z - \hat{\mathbf{f}}_z) \Lambda_z^{-1} (\mathbf{f}_z - \hat{\mathbf{f}}_z) \right] \quad (18)$$

with mean $\hat{\mathbf{f}}_z = \mathbf{L}_{21}^T \mathbf{L}_{22}^{-1} \mathbf{F}$ and variance $\Lambda_z = \Lambda_{11} - \mathbf{L}_{21}^T \mathbf{L}_{22}^{-1} \mathbf{L}_{21}$. Note that $\hat{\mathbf{f}}_z$ is simply a \mathbf{z} -dependent weighted linear combination of the measured data points.

Furthermore, assume that the related stochastic process, $\mathbf{f}_z^{\delta \mathbf{e}_i}$, where $\mathbf{f}_z^{\delta \mathbf{e}_i} = (\mathbf{f}_{(\mathbf{z} + \delta \mathbf{e}_i)} - \mathbf{f}_z) / \delta$ and \mathbf{e}_i is a unit basis vector, is well-defined in the limit as $\delta \rightarrow 0$, i.e. all the necessary probability distributions for a complete description exist. Denote the derivative stochastic process, i.e. the limiting random process, by $\mathbf{f}_z^{\mathbf{e}_i}$. The $E[\mathbf{f}_z^{\mathbf{e}_i}]$ as \mathbf{z} varies is interpreted as a fit to $\frac{\partial \mathbf{f}}{\partial z_i}(\mathbf{z})$ when the partial derivative of $\mathbf{f}(\mathbf{z})$ in the

direction \mathbf{e}_i exists. Provided the covariance $C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j})$ is sufficiently differentiable, it is known (O'Hagan 1992) that $\mathbf{f}_z^{\mathbf{e}_i}$ is itself Gaussian and that

$$E[\mathbf{f}_z^{\mathbf{e}_i}] = \frac{\partial}{\partial z_i} \mathbf{h}_f(\mathbf{z}) \quad ; \quad \mathbf{h}_f(\mathbf{z}) = E[\mathbf{f}_z] \quad (19)$$

where z denotes the i^{th} element of \mathbf{z} ; that is, the expected value of the derivative stochastic process is just the derivative of the expected value of the stochastic process. In addition,

$$E[\mathbf{f}_{z_0}^{\mathbf{e}_i} \mathbf{f}_{z_1}^{\mathbf{e}_i}] = \nabla_i^T \nabla_j^T C_f(\mathbf{z}_0, \mathbf{z}_1); C_f(\mathbf{z}_0, \mathbf{z}_1) = E[\mathbf{f}_{z_0} \mathbf{f}_{z_1}] \quad (20)$$

where $\nabla_i^T Q(\mathbf{z}_0, \mathbf{z}_1)$ denotes the partial derivative of $Q(\mathbf{z}_0, \mathbf{z}_1)$ with respect to the i^{th} element of its first argument, *etc.*

The above procedure can be repeated to construct second derivative stochastic processes. The means and covariances can be determined by recursive application of (19) and (20).

In the examples discussed in this paper, a straightforward smoothness prior covariance function is used which ensures that outputs associated with nearby inputs should have higher covariance than more widely separated inputs; specifically,

$$C(\mathbf{f}_{z_i}, \mathbf{f}_{z_j}) = \gamma \exp \left[-\sum_k \left((z_i)_k - (z_j)_k \right)^2 / 2\alpha_k \right] + \beta \delta_{ij} \quad (21)$$

where $(z_i)_k$ denotes the k^{th} element of vector \mathbf{z}_i . The value of α_k characterises the rate of variation of the function in dimension k , thereby, estimating the relative smoothness of different input dimensions. The parameter β is the variance of the measurement noise, n , on the output. The hyperparameters $(\beta, \alpha_k, \gamma)$ are adapted to maximise the

likelihood $p(\mathbf{M} | (\beta, \alpha_k, \gamma))$. The covariance function, (13), is sufficiently smooth for the derivative and second derivative stochastic processes to be well-defined and the relations (11) and (12) to apply (O'Hagan 1992).

References

- BAI, E.W., 1998, An Optimal Two-Stage Identification Algorithm for Hammerstein-Wiener Systems. *Automatica*, **34**, pp333-338.
- BAI, E.W., 2002, A Blind Approach to the Hammerstein-Wiener Model Identification. *Automatica*, **38**, pp967-979.
- GREEN, P.J., SILVERMAN, B.W., 1994, *Nonparametric Regression and Generalised Linear Models*. (Chapman & Hall, London).
- HABER, R., UNBEHAUEN, H., 1990, Structure Identification of Nonlinear Dynamic Systems – A Survey of Input/Output Approaches. *Automatica*, **26**, pp651-677.
- O'HAGAN, A., 1978, On curve fitting and optimal design for regression. *J. Royal Stat Soc. B*, **40**, 1-42.
- WILLIAMS, C. K. I., 1998, Prediction with Gaussian Processes: From linear regression to linear prediction and beyond. *In Learning and Inference in Graphical Models* (M. I. Jordan, Ed.), Kluwer.

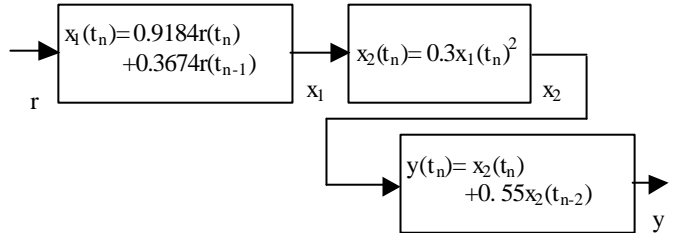


Figure 1 Block diagram representation of example system.