# On the Identification of Nonlinear Maps in a General Interconnected System [1]

Eric Wemhoff, Andrew Packard, and Kameshwar Poolla[†]
Department of Mechanical Engineering—University of California, Berkeley

**Abstract**

This paper is concerned with the problem of identifying static nonlinear maps in a general, structured interconnected system. These static nonlinear maps are *non-parametric* in that they do not have a natural parameterization that is known or suggested from an analytical understanding of the underlying process. Our technique involves selecting the nonlinear maps so as to maximize the "smoothness" or "staticness" of these maps while respecting the available input-output data and the noise model. These techniques avoid bias problems that arise when imposing artificial parameterizations on the nonlinearities. Computationally, these methods reduce to iterative least squares problems together with Kalman smoothing. Preliminary examples reveal the promise of these techniques.

## 1 Introduction

Nonlinear system identification is a challenging and important engineering problem. Much of the available literature treats nonlinear system identification in extreme generality, for example using Volterra kernel expansions, neural networks, or radial basis function expansions [1, 13, 14, 15]. These studies using nonlinear blackbox models typically offer asymptotic analyses, local convergence results, and general function approximation properties.

However a drawback of these methods is that the resulting models are ill-suited for many system analysis and control design problems. Typically we desire finite-dimensional state-description type models, and blackbox methods are unable to impose this constraint. Therefore it is our conviction that a completely general theory of nonlinear system identification can have little material impact on many practical problems that demand attention. We believe that it is more fruitful to study specific classes of nonlinear system identification problems, to devise appropriate algorithms for these, and to analyze the behavior of these algorithms. This experience can be collated together with experimental studies into a broader development of nonlinear system ID.

This paper is concerned with identification problems of interconnected structured nonlinear systems. The particular *a priori* knowledge concerning the individual elements in the interconnected system requiring identification is problem specific. For example some elements may be thought to be linear dynamic systems but with unknown state dimension or non-parametric linear systems (a noise model is one example of this), some may be dynamic systems parametrized by a few unknown quantities, and some may be static functions.

In particular we are concerned with problems in which there are non-parametric static nonlinear elements to be identified. Here by *non-parametric* we mean that these functions do not have a natural parameterization that is known or suggested from an analytical understanding of the underlying process. Such problems are particularly common in process control applications, or in nonlinear model reduction problems where an approximation is sought for a subsystem containing complex nonlinear dynamics.

Parameter estimation is a well-studied subject and its role in system identification has been thoroughly explored (see for example the texts [9, 18]). Non-parametric elements in an interconnected model are therefore often handled by coercing them to assume artificial parametrizations and invoking parameter estimation procedure. For example, static nonlinearities may be parameterized by neural networks, radial basis functions, wavelets, Volterra kernels, polynomial expansions, or (finite) Fourier series [4, 5, 14]. This includes many methods for Hammerstein and Wiener models, which are among the few classes of structured nonlinear model identification problems that have been studied significantly. [2, 3, 6, 7, 8].

We submit that this technique of coercing arbitrary "parameterizations" on the inherently non-parametric elements in an interconnected model and employing parameter estimation routines has several shortcomings. One is the vast number of parameters often needed—numerical procedures can become intractable with respect to speed and convergence considerations. The cost functions involved also become very complex, which brings up the

[1]Supported in part by an NSF Graduate Fellowship, by NASA-Dryden under Contract NAG4-124, and by NSF under Grant ECS 95-09539.
[†] Corresponding author. Department of Mechanical Engineering, University of California, Berkeley, CA 94720, Tel. (510) 642-4642, Email: poolla@jagger.me.berkeley.edu

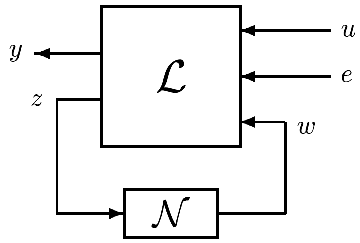


**Figure 1:** Model Structure

issue of choosing initial parameter estimates and basis functions. Poor basis function and/or initial parameter choice may likely result in convergence to false local minima, or worse, divergence, in the associated parameter estimation problem. Even with reasonable choices of basis functions, we may still incur a *bias* in the estimate of the nonlinearity due to undermodeling.

## 2 Identification of Static Nonlinearities

Consider the model structure shown in Figure 1. This model consists of a known linear part $\mathcal{L}$, and nonlinearities collected into a *static* block $\mathcal{N}$. The signal $u$ represents the known input, $e$ the unknown noise signal, and $y$ the measured output of the system. In this diagram $z$ and $w$ are signals internal to the model which represent the connection to the static nonlinearity and are not assumed to be available. We are interested in the case that non-parametric estimates are required for some or all of the components of $\mathcal{N}$.

Note that many nonlinear systems (and associated identification problems) can be rearranged to this LFT structure. This allows us to explore specific classes of structured nonlinear system identification problems in a common framework. Note also that the nonlinear block $\mathcal{N}$ will often have a block-diagonal structure, owing to the fact that nonlinearities appear in specific components in our interconnection.

Our approach is the following. Using a given input-output data record $(u, y)$ we seek to estimate the associated signals $(e, w, z)$. Our search will be guided by three desirable properties for these estimates. **P1**: the signals $(u, y, e, w, z)$ should be consistent with $\mathcal{L}$. **P2**: graphs from the (vector-valued) signals $z$ to $w$ should appear to be *static* nonlinear maps and agree with the structure of the block $\mathcal{N}$. **P3**: we should insist that the estimated signal $e$ be in some sense congruous with any *a priori* statistical information we may have regarding the noise process, that is, $e$ is a likely sample path. Amalgamating these requirements into a computationally attractive optimization problem is at the heart of our algorithms.

Partition $\mathcal{L}$ conformably as

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{yu} & \mathcal{L}_{ye} & \mathcal{L}_{yw} \\ \mathcal{L}_{zu} & \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix}.$$

We can write the set of all signals $(e, w)$ consistent with the input-output data $(u, y)$ as

$$\begin{bmatrix} e \\ w \end{bmatrix} = \begin{bmatrix} e^\circ \\ w^\circ \end{bmatrix} + \begin{bmatrix} \mathcal{B}_e \\ \mathcal{B}_w \end{bmatrix} f \tag{1}$$

where $f$ is a free signal. Here $\mathcal{B}$ is a basis for the null space of $\begin{bmatrix} \mathcal{L}_{ye} & \mathcal{L}_{yw} \end{bmatrix}$, and $(e^\circ, w^\circ)$ is a particular solution. Note also that the resulting signal $z$ which is consistent with these signals can be readily parameterized by $f$ as

$$z = z^\circ + (\mathcal{L}_{ze}\mathcal{B}_e + \mathcal{L}_{zw}\mathcal{B}_w) f. \tag{2}$$

All $(e, w, z)$ given by (1), (2) satisfy **P1**. We now develop a cost criterion that allows us to navigate the tradeoff between **P2** and **P3** as we search over $f$. As a beginning, let us restrict our attention to the case that $w$ and $z$ have the same size $r$, and the nonlinear block $\mathcal{N}$ has diagonal structure in that the i'th component $w_i$ of $w$ depends only on the i'th component $z_i$ of $z$. Thus $\mathcal{N}$ consists of $r$ single-input single-output static functions $\mathcal{N}_i$ such that $w_i(t) = (\mathcal{N}_i z_i)(t)\ \forall i, t$. Let $\nabla$ be the difference operator with action $(\nabla(u))(t) = u(t+1) - u(t)$. Also for a scalar real-valued sequence $q$, let $\mathcal{S}_q$ denote the permutation operator (linear) that sorts the values of $q$ in ascending order.

Now suppose we are given candidate signals $(z, w)$ that are allegedly related by $\mathcal{N}$. Following **P2**, we wish to develop a measure of "staticness" or smoothness of the static maps obtained by interpolating these signals. By smoothness we mean not the usual technical definition in terms of differentiability, but instead an intuitive notion that the graph of a static function is not too jagged up close. Since $\mathcal{N}_i$ is static and the operator $S_{z_i}$ only conducts a permutation, it follows that $\mathcal{N}_i(S_{z_i} z_i) = S_{z_i} w_i$. We suggest that the *dispersion* cost function defined by

$$\begin{aligned} D(z, w) &= \sum_{i=1}^{r} \sum_{t=0}^{L} \left\{ (\nabla S_{z_i} z_i)^2(t) + (\nabla S_{z_i} w_i)^2(t) \right\} \\ &= \sum_{i=1}^{r} \left\{ \|\nabla S_{z_i} z_i\|^2 + \|\nabla S_{z_i} w_i\|^2 \right\} \end{aligned} \tag{3}$$

is an appropriate measure of smoothness/staticness of the relations from $z$ to $w$. In (3) we sum the squared total-variation of the graphs of the component nonlinearities determined by interpolating the data $(z, w)$.

Intuitively, what we are trying to accomplish is this. Given scalar signals $z$ and $w$, form a scatter plot by plotting $(z(t), w(t))\ \forall t$. We are thinking of the resulting plot as a partial graph of the estimated nonlinearity. Now moving left to right, add up (the squares of) the lengths between points on the plot. If the plot does not look smooth, in that it is "cloudy", then this number will be large. Thus to some extent this operation captures our notion of what a static function should look like, while at

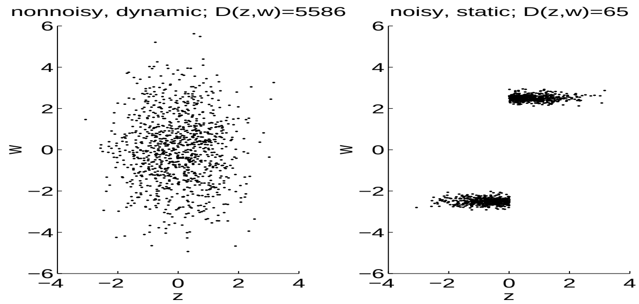


**Figure 2:** Illustrating the dispersion function

the same time being sufficiently simple to be practically useful.

In Figure 2 is a simple illustration of this idea. Here $z$ is a normal Gaussian white signal of length 1000. This drives a system and the output $w$ of the same length is plotted against $z$. In the left scatterplot the system is first order linear, time-invariant (LTI) with state space data $A = -.8, B = C = 1, D = 0$; on the right we have used the static nonlinearity $\{w(z) = 2.5$ if $z > 0$ and $-2.5$ else$\}$, with some added noise. Our native intuition immediately suggests that there is a static nonlinear input-output relationship prsent in the second data set. On the other hand, for the first data set we might reasonably conclude that there is no static function relating the input and output. The dispersion cost function reflects this, being 5586 in the first case and 65 in the second. Note that the dispersion function remains low even though this nonlinearity is not continuous.

Regarding property **P3**, it is often useful to assume that $e$ has the character of white and/or Gaussian noise. In this case we might demand of our estimates that they have similar characteristics. As a first cut we may use the size of $\|e\|$ as an indication of fitness.

Combining criteria **P2** and **P3** suggests the cost function

$$J(f) = \gamma \|e_f\|^2 + D\left(z_f, w_f\right) \tag{4}$$

(where now we show explicitly the dependence of the estimates on the free parameter $f$). The parameter $\gamma$ allows us to balance the two fitness measurements in identifying the nonlinear maps of $\mathcal{N}$. Experience with computational examples shows that a choice of $\gamma$ which produces a noise estimate with size comparable to the variance of the noise model is often a good choice in terms of accuracy of the estimates $e, w, z$. It is possible to locate this value of $\gamma$ iteratively.

The optimization problem we have assembled is thus

$$\min_f \left\{\gamma \|e_f\|^2 + \|\nabla S_{z_f} z_f\|^2 + \|\nabla S_{z_f} w_f\|^2\right\}, \tag{5}$$

where we write $\|\nabla S_x y\|^2$ to mean $\sum_{i=1}^r \|\nabla S_{x_i} y_i\|^2$. We should stress that this does not capture all the aspects we might wish to impose the estimates. Indeed, *whiteness* of the noise signal $e$ and requiring that $e$ and $u$ be uncorrelated (as in open-loop identification) are other possibilities which come to mind.

The problem (5) is, in general, a nonlinear programming problem. In the special case where the signal $z$ is measured (i.e. $[\mathcal{L}_{ze}\mathcal{L}_{zw}]\,\mathcal{B} = 0$) it simplifies considerably. In this case, the free signal $f$ does not affect the inputs $z$ to $\mathcal{N}$. Then, the cost criterion simplifies to

$$J(f) = \gamma \|e_f\|^2 + \sum_{i=1}^r \|\nabla S_{z_i} w_{f,i}\|^2$$

Since $(e_f, w_f)$ are affine in $f$, and the permutations $S_*$ are independent of $f$, this is a least squares problem.

## 3 Computational Issues

Solution of the central optimization problem (5) is nontrivial. The source of the difficulty is that the permutation operators $S_{z_f}$ are themselves dependent on $f$. However, if $S$ is fixed, the problem reduces to an instance of least squares. This suggests a bootstrapping algorithm to solve (5):

0 | initialize $f^{(0)}$
while stopcheck
  fix the permutation operators, $S = S_{z_{f^{(k)}}}$
1 | $f^{(k+1)} = \arg\min_f \left\{\gamma\|e_f\|^2 + \|\nabla S z_f\|^2 + \|\nabla S w_f\|^2\right\}$
  update $k = k + 1$
end

Note that Step 1 above is simply least squares. We would like to remark that in the event $z$ is measured, this bootstrapping procedure becomes unnecessary as the problem (5) reduces to just this step, since $S$ is fixed. Our initial experience indicates the promise of this technique. While it is unreasonable to expect global convergence, it appears possible to establish a local convergence result here using the methods of [10, 11].

The minimization problem of Step 1, which we capriciously refer to as "simply" least squares, requires serious attention. In the event the data record is of modest length, we may explicitly form the matrix representations of $\nabla, S, \mathcal{B}_w$, and $\mathcal{B}_e$ that appear in $e_f, w_f$ and then use a standard least squares solver. (For example the matrix representations of $\mathcal{B}_w$ and $\mathcal{B}_e$ can be found using singular value decomposition of the matrix representation of $[\mathcal{L}_{ye}\mathcal{L}_{yw}]$.)

More often, however, we are faced with having a sizeable input-output data record (greater than, say, 200 samples on a modern PC running `Matlab`). In this case, storage requirements render the direct method above impractical. We therefore suggest an iterative least squares procedure [16]. For the least squares problem

$$\min_x \|Ax - b\|$$

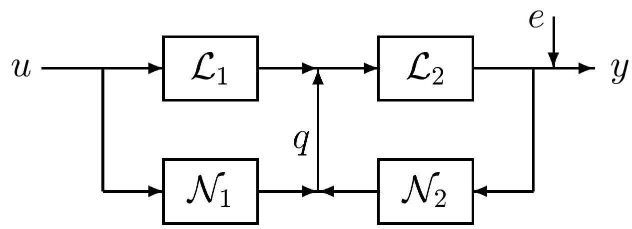


**Figure 3:** Example 1

with $A$ some linear operator, this iteration proceeds as

$$x^{(k+1)} = x^{(k)} - \frac{\|A^* r\|^2}{\|AA^* r\|^2} A^* r$$

where $r = Ax^{(k)} - b$. We therefore have to repeatedly compute the actions of $A$ and its adjoint operator $A^*$. In our situation $A$ is a product of various Toeplitz operators, permutation operators, and the difference operator $\nabla$. We can avoid forming large Toeplitz matrices by determining state-space realization of the complimentary inner factors of $[\mathcal{L}_{ye}\mathcal{L}_{ze}]$ (see [19] for details). Such spectral factorization methods are numerically necessary because $\mathcal{L}$ may be unstable and/or non-minimum phase. The adjoints of the Toeplitz operators involve (anti-causal) filtering. Adjoints of permutation operators are themselves permutations and can be rapidly found by a sorting procedure. The adjoint of $\nabla$ is again a (backward) difference operator. This iterative least squares procedure is employed in the first example of the next section.

These methods for solving Step 1 above can be slow. An interesting open problem is to exploit the special structure of our least squares problems (i.e. Toeplitz, difference, and permutation operator products) to devise a fast recursive least squares solver. The impediment here appears to be the permutation operators, without which the problem reduces to standard Kalman smoothing [20].

## 4 Examples

In this section we present two simulation examples.

### 4.1 Example 1

Consider the model of Figure 3. Here, the nonlinearities $\mathcal{N}_1$ and $\mathcal{N}_2$ are to be identified. The (possibly unstable) linear systems $L_1$ and $L_2$ are known. In this situation, it is easy to verify that we essentially have access to the signal $q = (\mathcal{N}_1 u + \mathcal{N}_2 y)$, modulo additive output noise. This follows because we can write $q = \mathcal{L}_2^{-1} y - \mathcal{L}_1 u$. The "inversion" here is to be conducted using Kalman smoothing methods as $\mathcal{L}_2$ need not be minimum-phase. From the signal $q$ we are to infer the individual nonlinear maps. Plotted in the panels above are the results of our technique as applied to this example. The solid lines are the graphs of the true nonlinearities. The dots are the estimates based on 2000 samples of input-output data using a white-noise input and a bandpass multi-tone input. In both cases, the data was corrupted by additive output noise at a signal-to-noise ratio of $15db$.

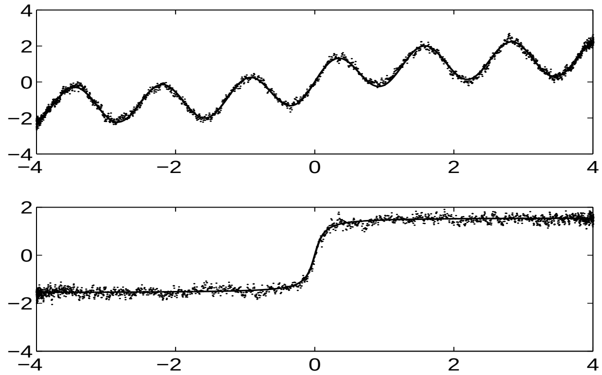


**Figure 4:** Example 1: White-noise input

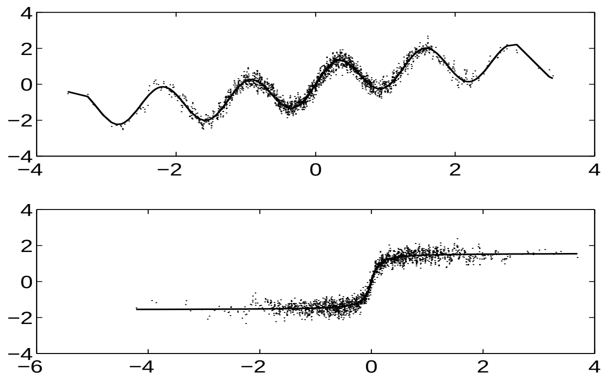


**Figure 5:** Example 1: Multi-tone input

The noise is not, however, incorporated in the modeled structure. Here, we used the iterative least squares procedure (see Section 4), for which we found 200 iterations to be adequate. The size of our input-output data record compelled us to use the iterative method.

### 4.2 Example 2

In our second example, we have as in Figure 1 a single single-input single-output static nonlinearity in feedback around a known linear system with 3 states, scalar input $u$ and noise input $e$, and scalar measured output $y$. The particular state-space data for the example is:

$$\begin{bmatrix} A & B_u & B_e & B_w \\ C_y & D_{yu} & D_{ye} & D_{yw} \\ C_z & D_{zu} & D_{ze} & D_{zw} \end{bmatrix} =$$

$$\begin{bmatrix} 0.4221 & 0.7440 & 0.9682 & 0.5686 & 0 & -3.5047 \\ -2.3878 & 5.6667 & 6.1124 & -0.2872 & 0 & -2.6991 \\ 2.4575 & -5.0701 & -5.4888 & -1.2706 & 0 & 0.9049 \\ 0.5442 & -0.3254 & -0.2325 & 0 & 0.2000 & 0 \\ 2.1961 & -3.8272 & -4.4564 & 0 & 0 & 0 \end{bmatrix}$$

Initial state is zero, both inputs are are white noise. The noise $e$ is output additive, with a 15dB signal to noise ratio. A 200 sample data record was generated with no undermodelling, and we proceeded to estimate the unknown signals following 5. We did not employ iterative least squares methods, but instead chose to form the permutation matrices and invoke the MATLAB solver. This was possible given the modest size of our input-output data record.

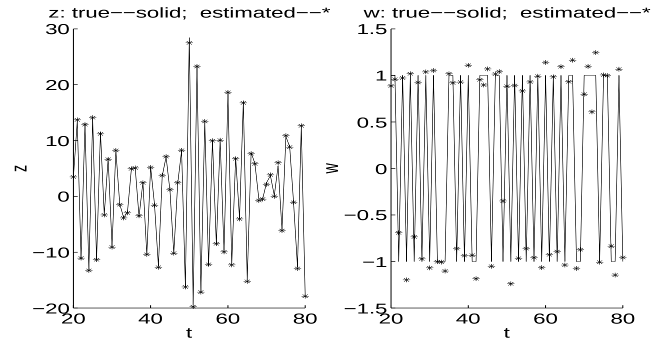


**Figure 6:** Example 2: Comparing estimates of $z$, $w$ to the true signals for $\gamma = 2$.

We obtained estimates for 2 values of the design variable $\gamma$. After trying several values, we found that with $\gamma = 0.2$ the estimate of $e$ has size close to that of the true signal. This same value produces good estimates of $z$, $w$. A portion of these estimates along with the true simulation signals are shown in Figure 6. See Figure 8 for a similar plot for $e$. A scatterplot of the estimates for $w$, $z$ is made in Figure 7–this is the estimate of the nonlinearity.

We also show results for $\gamma = 2$ in Figures 7 and 8. Referring to the optimization procedure 5, note that a higher value here tends to penalize the size of $e$ more. Thus the resulting estimate is smaller. Note that the scatter plot becomes more scattered over the $\gamma = 0.2$ case. Thinking intuitively, this is due to the fact that the optimization would rather explain the (noisy) $u, y$ data using more $w$ and less $e$; hence the $w$ estimate looks more like noise.

The values of the cost functions $J(e, z, w)$ and $D(z, w)$ for certain signals are of interest. Take the first value for $\gamma$, 0.2. For the true simulation signals $J = 47.6$ and $D = 39.5$. At the estimates we found for these signals, $J = 37.7$ and $D = 30.0$. Picking a random value for the free signal $f$, we typically find $J = 74$ and $D = 60$. Thus we see that the iteration described in Sections 2 and 3 have managed to produce estimates which lower the cost function.

## 5 Semi-parametric System Identification

To this point we have described a possible method for non-parametric identification of static nonlinearities interconnected with a known linear system as in Fig. 1. A more general situation is shown in Fig. 9 in which now the linear system is not completely known, but has unknown parameters. We also include here a separate LTI noise model system $\mathcal{W}$ for which a non-parametric approximation may be desired.

For this structure, we suggest a combined parametric/non-parametric approach that retains the advantages of each method. The basic structure of our proposed approach is shown below. We begin with an input-output data record $(u, y)$. Then,

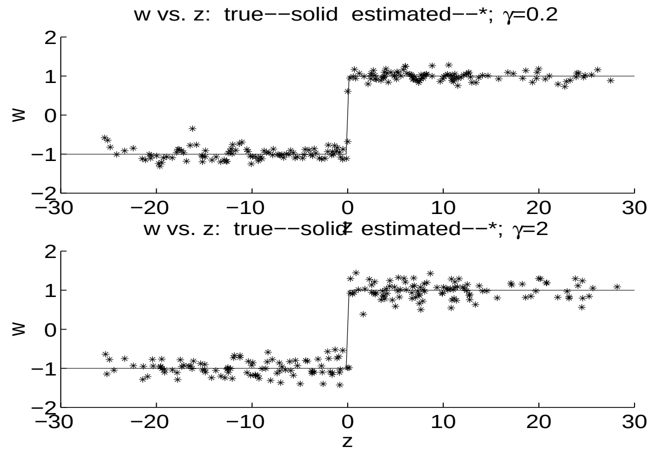


**Figure 7:** Example 2: Scatterplot, $\hat{w}$ vs. $\hat{z}$, for 2 $\gamma$.

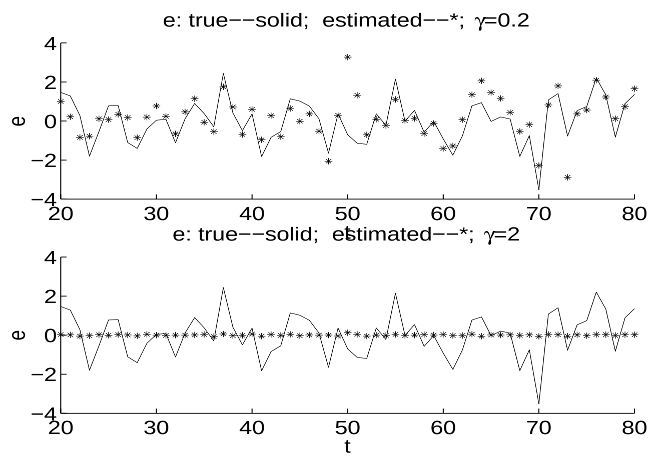


**Figure 8:** Example 2: Comparing the estimate of $e$ to the true noise signal, for 2 $\gamma$.

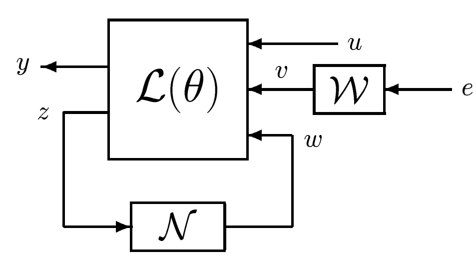


**Figure 9:** Model Structure

```
1 | initialize N
  | while stopcheck
2 |    estimate θ, W
  |
3 |    estimate e, w, z
  |
4 |    update W
  |    update N
  | end
```

Step 1 is delicate, and will generally require the use of *a priori* problem-specific information supplied by the user. We may then use classical methods for linear time-invariant system identification [9] to make estimates of $\theta$ and $\mathcal{W}$. The estimation of $\theta$ in Step 2 may be conducted using classical parameter estimation techniques such as (linearized) maximum likelihood estimation [12]. another possibility when $\mathcal{L}$ is not well known is to use data from small-signal experiments so as not to excite the nonlinear dynamics. Step 3 is the focus of this paper, and is discussed above. In Step 4, the noise model $\mathcal{W}$ can be updated by computing the correlation of the estimated noise $e$, and realizing its power spectral density. This is a classical problem in signal processing [17]. The nonlinearities $\mathcal{N}$ may be revised graphically using a scatter-plot of $\hat{w}$ versus $\hat{z}$, by using a smoothed table look-up curve fit, or by standard spline interpolation. These methods respect the non-parametric nature of the nonlinearities $\mathcal{N}$ and the noise model $\mathcal{W}$.

## 6 Conclusions

We have presented a novel method for identification of static nonlinearities embedded in a general interconnected system with known structure and linear components. The appeal of this method is twofold. First, it combines a smoothness criteria (to reflect our desire that the estimated nonlinearity be static) with a noise penalty term. Second, we have offered computationally tractable methods for optimizing this cost functional. We have also described how our method fits into a larger framework for identification of general interconnected nonlinear systems.

In the future, we need to develop and implement more sophisticated computational algorithms for these methods, as well as address convergence issues. Further, a consideration of identifiability and experiment design is essential. Finally, we wish to further address the situation when the linear system is unknown.

## References


[1] S. A. Billings, S. Y. Fakhouri, "Identification of a Class of Non-linear Systems Using Correlation Analysis," *Proc. IEE,* v.125, pp.691-7 (1978).

[2] M. Boutayeb, M. Darouach, H. Rafaralahy, G. Krzakala, "A New Technique for Identification of MISO Hammerstein Model," *Proc. 1993 ACC,* San Francisco, CA, v.2, pp.1991-2 (Jun. 1993).

[3] C. Chen, S. Fassois, "Maximum likelihood identification of stochastic Wiener-Hammerstein-type non-linear systems," *Mech. Syst. Sig. Proc.*, v.6, n.2, pp.135-53 (Mar. 1992).

[4] W. Greblicki, "Non-parametric orthogonal series identification of Hammerstein systems," *Int. J. Syst. Sci.*, v.20, n.12, pp.2355-67 (Dec. 1989).

[5] A. Kryzak, "Identification of discrete Hammerstein systems by the Fourier series regression estimate," *Int. J. Syst. Sci.*, v.20, n.9, pp.1729-44 (Sept. 1989).

[6] K. S. Narendra, P. G. Gallman, "An Iterative Method for the Identification of Nonlinear Systems Using the Hammerstein Model," *IEEE Trans. Auto. Contr.*, v.11, n.7, pp.546-50, (Jul. 1966).

[7] M. Pawlak, "On the series expansion approach to the identification of Hammerstein systems," *IEEE Trans. Auto. Contr.*, v.36, n.6, pp.763-7 (Jun. 1991).

[8] P. Stoica, "On the Convergence of an Iterative Algorithm Used for Hammerstein System Identification," *IEEE Trans. Auto. Contr.*, v.26, n.4, pp.967-69 (Aug. 1981).

[9] L. Ljung, *System Identification, Theory for the User*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[10] J.E. Dennis, Jr. and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall, 1983.

[11] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*, Academic Press, 1981.

[12] G. Wolodkin, S. Rangan, and K. Poolla, "An LFT approach to parameter estimation," *Proc. of the Automatic Control Conference*, San Francisco, pp. 2088-2092, June 1997.

[13] T.A. Johansen, "Identification of Non-linear Systems using Empirical Data and Prior Knowledge–An Optimization Approach," *Automatica*, March 1997, vol.32 (no.3), pp.337–56.

[14] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsdy, "Nonlinear Black-box Modeling in System Identification: a Unified Overview," *Automatica*, 1995, vol.31 (no.12), pp.1691–1724.

[15] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, and Q Zhang, "Nonlinear Black-box Models in System Identification: Mathematical Foundations," *Automatica*, 1995, vol.31 (no.12), pp.1752–1750.

[16] G. Golub and C.F. Van Loan, *Matrix Computations*, 2nd edition, The Johns Hopkins University Press, 1989.

[17] A.V. Oppenheim and R.W. Schafer, *Discrete-time signal processing*, Prentice-Hall, 1989.

[18] R. Deutsch, *Estimation Theory*, Prentice-Hall, 1965.

[19] B. Francis, *A course in $H_\infty$ control theory*, Springer-Verlag, 1987.

[20] M.S. Grewal, *Kalman filtering : theory and practice*, Prentice-Hall, 1993.