

Inverse Regression for the Wiener Class of Systems

Christian Lyzell, Martin Enqvist

Division of Automatic Control

E-mail: lyzell@isy.liu.se, maren@isy.liu.se

14th November 2011

Report no.: LiTH-isy-R-3032

Submitted to 16th IFAC Symposium on System Identification

Address:

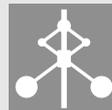
Department of Electrical Engineering

Linköpings universitet

SE-581 83 Linköping, Sweden

WWW: <http://www.control.isy.liu.se>

AUTOMATIC CONTROL
REGLERTEKNIK
LINKÖPINGS UNIVERSITET



Abstract

The concept of inverse regression has turned out to be quite useful for dimension reduction in regression analysis problems. Using methods like *sliced inverse regression* (SIR) and *directional regression* (DR), some high-dimensional nonlinear regression problems can be turned into more tractable low-dimensional problems. Here, the usefulness of inverse regression for identification of nonlinear dynamical systems will be discussed. In particular, it will be shown that the inverse regression methods can be used for identification of systems of the Wiener class, that is, systems consisting of a number of parallel linear subsystems followed by a static multiple-input single-output nonlinearity. For a particular class of input signals, including Gaussian signals, the inverse regression approach makes it possible to estimate the linear subsystems without knowing or estimating the nonlinearity.

Keywords: System identification; Dimension reduction; Inverse regression.

Inverse Regression for the Wiener Class of Systems

Christian Lyzell, Martin Enqvist

2011-11-14

Abstract

The concept of inverse regression has turned out to be quite useful for dimension reduction in regression analysis problems. Using methods like *sliced inverse regression* (SIR) and *directional regression* (DR), some high-dimensional nonlinear regression problems can be turned into more tractable low-dimensional problems. Here, the usefulness of inverse regression for identification of nonlinear dynamical systems will be discussed. In particular, it will be shown that the inverse regression methods can be used for identification of systems of the Wiener class, that is, systems consisting of a number of parallel linear subsystems followed by a static multiple-input single-output nonlinearity. For a particular class of input signals, including Gaussian signals, the inverse regression approach makes it possible to estimate the linear subsystems without knowing or estimating the nonlinearity.

1 Introduction

Identification of a nonlinear system is a challenging task in the absence of prior information about the system structure. Although it is possible to propose and estimate a general high-order model, a large dataset is typically needed to get a robust result. With a limited amount of data, a common strategy is to restrict the class of models by imposing some structural constraints [Ljung, 1999, Pearson, 1999]. This approach might work well if these constraints actually reflect some properties of the true system. However, if the structural mismatch between the model and the system is too large, the resulting model will only be a coarse approximation with limited usefulness.

This problem can be handled by a thorough model validation and a choice of a new model structure in case the original one does not pass the validation step. Obviously, this can be a time-consuming and user-dependent strategy. A more convenient and automatic alternative would be to use a model structure that is general enough to be able to approximate a large class of systems together with an identification method that automatically can detect and exploit some useful system structure in data. Such an approach will be described here for the *Wiener class of systems* [see, for example, Schetzen, 1980, 2010], using statistical methods for dimension reduction based on *inverse regression* for the structure detection and exploitation.

The Wiener class of systems is an important class of nonlinear dynamical systems, both from a theoretical and a practical point of view. This system structure is

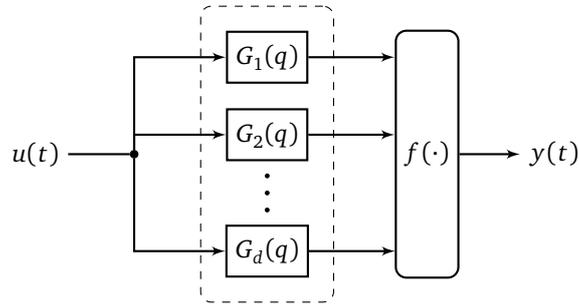


Figure 1: An illustration of the Wiener class of systems, where the input enters a finite number of linear time-invariant systems followed by a static nonlinearity.

illustrated in Figure 1 and consists of a finite number of parallel linear subsystems followed by a static multiple-input single-output nonlinearity. It can be shown, under fairly weak assumptions on the input signal $u(t)$, that all systems with *fading memory* can be approximated arbitrarily well by the Wiener class of systems [Boyd and Chua, 1985].

The identification of the Wiener class of systems in the case of a single branch ($d = 1$) and the case when the structure of the Wiener system is known has undergone extensive research and a wide variety of methods exist [see, for instance, Giri and Bai, 2010, for an overview]. This paper focuses on the multiple branch case, where the number of branches is unknown and the linear subsystems have a *finite impulse response* (FIR).

A similar problem in the statistical literature is the special case of nonlinear regression when the model can be written

$$y(t) = f(B^T \varphi(t), e(t)), \quad B \in \mathbb{R}^{n_b \times d}, \quad (1)$$

where $\varphi(t)$ contains the known regressor variables, $e(t)$ is an unknown noise component, and t is an arbitrary index variable in the general case, but time here. If d is significantly smaller than n_b , this estimation problem is much more tractable than the general problem of finding an unstructured nonlinear function of all n_b components of $\varphi(t)$. When considering the inference problem of estimating the B matrix in (1), given data, it is natural to think in terms of forward regression, that is, finding a suitable projection via the minimization of some cost function. Typically, this approach involves parameterizing the nonlinearity in some eligible manner [see, for example, Friedman and Stuetzle, 1981]. One drawback is that the corresponding optimization problem often becomes nonconvex, with several local minima, and that an initial estimate is needed to start the local search algorithm. Other drawbacks include the difficulty of estimating the dimension of the corresponding subspace and that the optimization problem may be computationally expensive to solve. For instance, the cost function might have to be minimized over the Grassmann or Stiefel manifolds [Friedman, 1987].

However, the nonlinear regression approach based on (1) can be simplified significantly if B can be estimated *before* f , since this enables a strategy where the regressors first are projected onto a d -dimensional subspace followed by a nonlinear regression on this lower-dimensional space. Estimation of B for large classes of unknown functions f turns out to be possible using an inverse regression approach. This has been a very active research topic within the statistical community

during the last two decades and a number of methods are now available, for example, *sliced inverse regression* (SIR) [Li, 1991], *principal Hessian directions* (PHD) [Li, 1992], and *directional regression* (DR) [Li and Wang, 2007]. Although there is a theoretical connection with *Bussgang's theorem* [Bussgang, 1952] and the results about *separable signals* [Nuttall, 1958], which have been used extensively for system identification, the inverse regression approach has to the authors' knowledge not been applied to identification of dynamical systems with colored inputs. This is the main topic that will be discussed here.

This paper is organized as follows. In Section 2, a brief review of the basic theoretical results of some inverse regression methods is given. The statistical inference is described in Section 3 and followed by a simulation study in Section 4, where the performance of the proposed methods for the identification of Wiener class of systems is analyzed. The paper is then summarized and concluded in Section 5.

2 Inverse regression

The basic idea in inverse regression [similarly to Bussgang, 1952] is to limit the distribution of the input signal. The most commonly used assumption is known as the *linear design condition*.

Assumption 1. For any $a \in \mathbb{R}^{n_b}$, there exist $c_0 \in \mathbb{R}$ and $c \in \mathbb{R}^d$ such that

$$\mathbb{E}(a^\top \varphi(t) | B^\top \varphi(t)) = c_0 + c^\top B^\top \varphi(t). \quad (2)$$

It can be shown [Cook and Weisberg, 1991] that the above condition is fulfilled, for instance, when the regressors are elliptically distributed, that is, when the level curves of the corresponding probability density function are ellipsoids [see, for example, Eaton, 1986]. The full implications of the assumption are not yet known and a more detailed discussion can be found in Li and Wang [2007]. Now we are ready to state the fundamental theorem for the inverse regression methods which are based on first order moments.

Theorem 1 (Li [1991]). *If the system is given by (1) and Assumption 1 is fulfilled, it holds that*

$$\mathbb{E}(\varphi(t) | y(t)) - \mathbb{E}(\varphi(t)) \in \mathcal{R}(\text{Cov}(\varphi(t))B), \quad (3)$$

where the operator \mathcal{R} returns the column space of its argument.

In the following, it will be assumed that the regressors are *standardized*, that is, the regressors are transformed to have zero mean and an identity covariance matrix,

$$\zeta(t) \triangleq \Gamma^{-1}(\varphi(t) - \mathbb{E}(\varphi(t))), \quad (4)$$

where we have defined $\Gamma \triangleq \text{Cov}(\varphi(t))^{1/2}$ and (3) can now be written as

$$\mathbb{E}(\zeta(t) | y(t)) \in \mathcal{R}(\Gamma B). \quad (5)$$

This quite interesting result shows that the only information needed to obtain an approximation of the column space of B is the conditional expectation $\mathbb{E}(\zeta(t) | y(t))$. This entity can be estimated in a number of different ways, for instance, parametrically as in [Bura and Cook, 2001] or nonparametrically as in [Li, 1991, Bura,

2003]. In this paper we will make use of the latter methods, which will be referred to as *sliced* and *local inverse regression*, abbreviated SIR and LIR, respectively.

One of the fundamental limitations of the usability of the result given in Theorem 1 is that the nonlinearity appearing in (1) is not allowed to be even. If this is the case, then $E(\zeta(t)|y(t))$ is zero and no information concerning the column space of B is present. To remedy this limitation, several methods based purely on second order moments have been suggested in the literature [see, for example, Cook and Weisberg, 1991, Li, 1992] and more recently a combination of both moments [Li et al., 2005, Li and Wang, 2007]. In addition to Assumption 1, these methods are also based on the following condition.

Assumption 2. For any $a \in \mathbb{R}^{n_b}$, it holds that

$$\text{Var}(a^T \varphi(t) | B^T \varphi(t)),$$

is a nonrandom number.

In this paper, the use of the *directional regression* (DR) method is considered, which has shown to be advantageous in terms of accuracy when compared to other moment based inverse regression methods [Li and Wang, 2007]. Now, denote by $\zeta(t)$ the standardized version of $\varphi(t)$ according to (4). If we define

$$A(y(s), y(t)) \triangleq E((\zeta(s) - \zeta(t))(\zeta(s) - \zeta(t))^T | y(s), y(t)),$$

then the following theorem holds.

Theorem 2 (Li and Wang [2007]). *Let the system be given by (1) and let Assumption 1 and 2 be fulfilled. If s and t are such that $(\zeta(s), y(s))$ and $(\zeta(t), y(t))$ are independent, then $2I - A(s, t) \in \mathcal{R}(\Gamma B)$, where I is the identity matrix of an appropriate size.*

In the theorem above, independence between different points in time is needed, which indicates that DR may not be a natural choice for the identification of dynamical systems where the signals may contain dynamical dependencies. This possible issue will be investigated in this paper and analyzed in future research. Now, let us turn our attention to the implementation of the above methods.

3 Statistical Inference

The use of the results in the previous section for the inference about $\mathcal{R}(B)$ is straightforward and the methods only differ, as we shall see, in which of the quantities in Theorem 1 and 2 that is used and how it is estimated. In the following, consider that a dataset $(\varphi(t), y(t))_{t=0}^{N-1}$ is given in accordance to (1) and

$$\varphi(t) \triangleq \begin{pmatrix} u(t) & u(t-1) & \cdots & u(t-n_b+1) \end{pmatrix}^T. \quad (6)$$

Algorithm 1 (Inverse regression). Given a dataset, an estimate \hat{T}_B of a basis for $\mathcal{R}(B)$ is returned.

1. Standardize the regressors

$$\zeta(t) \triangleq \hat{\Sigma}^{-1/2}(\varphi(t) - \hat{\mu}),$$

where $\hat{\mu}$ and $\hat{\Sigma}$ is the sample mean and covariance, respectively.

2. Construct the matrix M for a method of choice, see Sections 3.1-3.3 below.
3. Find the d leading singular values of M and the corresponding left singular vectors U_d .
4. Then $\hat{T}_B = \hat{\Sigma}^{-1/2}U_d$ is an estimate of a basis for $\mathcal{R}(B)$.

In the algorithm above, the only method dependent part is Step 2. In the following sections, three different methods for constructing M are given corresponding to SIR, LIR and DR, respectively. The first two methods are based on the results in Theorem 1 while the third method relates to Theorem 2. In the Step 3 of Algorithm 1, one needs to determine an appropriate value for the dimension d . This can be done in several ways, either by the use of some test statistic [Bura and Yang, 2011] or via a straightforward inspection of a plot of the eigenvalues.

3.1 Sliced Inverse Regression

The original work of Li [1991] uses a rather simple estimator for the conditional expectation appearing in (5) and is based on a discretized version of the output. The discretization uses a partition

$$\{\mathcal{Y}_j, j = 1, 2, \dots, J\},$$

of the support of $y(t)$, where the number J is supplied by the user. There are many different ways to make such a partition of the data. Here, we are going to use a simple data dependent partitioning scheme. Assume that the output has been sorted in ascending order, that is, $y(s) \leq y(t)$ for all $s \leq t$, and that $\zeta(t)$ has been rearranged accordingly. Let k be the largest integer such that $kJ \leq N$. Then

$$\mathcal{Y}_j \triangleq \{y(t), (j-1)k \leq t < jk\}, \quad j = 1, 2, \dots, J-1, \quad (7)$$

and $\mathcal{Y}_J \triangleq \{y(t), (J-1)k \leq t \leq N-1\}$ is a partition of the output data. With this partitioning scheme, the conditional expectation $E(\zeta(t) | y(t) \in \mathcal{Y}_j)$ can be estimated by

$$\hat{\mu}_j \triangleq \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} \zeta(i), \quad (8)$$

where \mathcal{I}_j is the set of indices belonging to slice j and $|\mathcal{I}_j|$ the number of such elements. The weighted estimator of the covariance matrix of (8) can now be determined by

$$M_{\text{SIR}} \triangleq \frac{1}{N} \sum_{j=1}^J |\mathcal{I}_j| \hat{\mu}_j \hat{\mu}_j^T, \quad (9)$$

and used in Step 2 of Algorithm 1 to yield the SIR estimate of a basis of $\mathcal{R}(B)$. The SIR estimator is illustrated for a simple example in Figure 2.

The consistency of the SIR estimator is shown in Li [1991] for the case of independent identically distributed regressors and generalized to allow finite dependence in Lyzell and Enqvist [2011]. A MATLAB implementation is freely available via Cook et al. [2011].

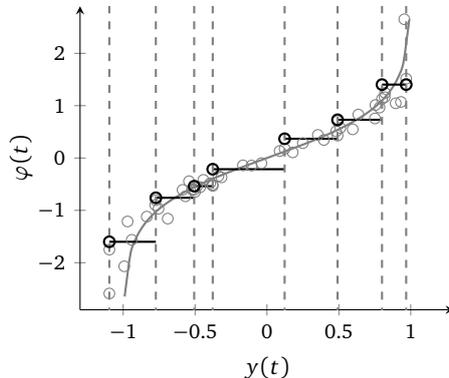


Figure 2: A dataset $(\varphi(t), y(t))_{t=0}^{48}$ has been generated from $y(t) = \tanh(\varphi(t)) + v(t)$, where $\varphi(t)$ is Gaussian with zero mean and unit variance and $v(t)$ is white Gaussian noise with zero mean and variance 0.01. The measurements are depicted by the gray circles while the gray solid line shows the noise free function values. The partition explained in Section 3.1, illustrated with the dashed vertical lines, yields the estimate of the conditional expectation illustrated by the black piecewise constant function.

3.2 Local Inverse Regression

The SIR estimator explained above uses equal weight for all observations within a slice when estimating the conditional expectation in (5). A more reasonable approach, especially when the observations are far apart, as might be the case when the number of available observations is small, is to let the weights decrease with the distance. Furthermore, SIR only fits a constant to data and the performance of the estimator might improve by increasing the order and adapt a linear model. In addition, the partitioning of the output can become quite complicated in higher dimensions. This reasoning leads us to examine local linear models [see, for instance, Fan and Gijbels, 1996], which is a special case of the linear smoother that have previously been considered in the framework developed by Bura [2003].

The basic idea is to estimate the conditional expectation (5) for different values of $y(t)$. In the following, let $\tilde{y}(t) = y(t) - \bar{y}$ for some value \bar{y} . A local linear model around \bar{y} is then determined by

$$\hat{\beta} = \arg \min_{\beta} \sum_{t=0}^{N-1} K_h(\tilde{y}(t)) \|\zeta(t) - \beta_0 - \beta_1 \tilde{y}(t)\|_2^2, \quad (10)$$

where K_h is some function which is positive and tends to zero as the argument increases. It holds that

$$\mathbb{E}(\varphi(t) | y(t) = \bar{y}) \approx \hat{\beta}_0, \quad (11)$$

and we let M_{LIR} in Algorithm 1 have columns corresponding to (11) for different measurements \bar{y} . This is a slightly different choice compared to Bura [2003] but seems to improve the accuracy of LIR to some extent.

In this paper, the *Epanechnikov* kernel

$$K(u) \triangleq \frac{3}{4} \max(1 - u^2, 0), \quad K_h(u) \triangleq K(u/h)/h,$$

which is nonzero only on a finite interval, is chosen with a kernel width h corresponding to the k nearest neighbors, that is, h is chosen as the k th smallest value of $|\tilde{y}(t)|$. Further information about local polynomial modeling can be found in, for example, Fan and Gijbels [1996].

3.3 Directional Regression

The DR method uses the result of Theorem 2 and the partitioning scheme described in (7) for inference. It can be shown [Li and Wang, 2007], under some independence assumptions, that the expected value of matrix valued stochastic variable $(2I - A(s, t))^2$ appearing in Theorem 2 can be estimated consistently by

$$M_{\text{DR}} \triangleq \frac{2}{N} \sum_{j=1}^J |\mathcal{I}_j| \hat{\Sigma}_j^2 + 2M_{\text{SIR}}^2 + 2 \left(\frac{1}{N} \sum_{j=1}^J |\mathcal{I}_j| \hat{\mu}_j^T \hat{\mu}_j \right) M_{\text{SIR}} - 2I, \quad (12)$$

where all the terms are defined in Section 3.1, except for

$$\hat{\Sigma}_j \triangleq \frac{1}{|\mathcal{I}_j|} \sum_{i \in \mathcal{I}_j} \zeta(i) \zeta(i)^T. \quad (13)$$

A MATLAB implementation is given in the package Cook et al. [2011] and further information concerning the derivation (12) can be found in Li and Wang [2007].

4 Simulations

In this section, the performance of the inverse regression methods presented above will be evaluated on simulated data from systems of the Wiener class. The first simulation study concerns the single branch case where the effects of different input signals and nonlinearities are analyzed. In the second simulation study, an unknown number of branches is present and the ability to indicate the number of linear subsystems correctly is evaluated.

As a measure of quality of an estimate the *angle between subspaces* will be used [see, for instance, Li and Wang, 2007]. Let A be an estimate of B . The angle between the column spaces of A and B is defined by

$$\text{angle}(A, B) \triangleq \arcsin(\|\Pi_A - \Pi_B\|_2), \quad (14)$$

where Π_A denotes the orthogonal projection on the column space of A . This generalizes the angle between vectors and can be determined, for example, in MATLAB with the command `subspace`.

The quality of the data will be measured in terms of the *signal to noise ratio* defined here by

$$\text{SNR} = \frac{\text{Var}(y_{\text{nf}}(t))}{\text{Var}(y(t) - y_{\text{nf}}(t))}, \quad (15)$$

where $y_{\text{nf}}(t)$ denotes the noise-free part of $y(t)$.

In the following, let $\mathcal{U}(a, b)$ denote the uniform distribution on the interval (a, b) and let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2 .

4.1 A single branch Wiener system

Consider the system given by

$$y(t) = f(B^T \varphi(t)) + v(t), \quad v(t) \sim \mathcal{N}(0, \sigma_v^2), \quad (16)$$

where $B \in \mathbb{R}^{13}$ is chosen as

$$B_i = \begin{cases} 3/2^i, & i = 4, 5, \dots, 10, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

The variance σ_v^2 of the additive noise will be chosen later to yield a specific SNR as defined in (15). Furthermore, the effects of the choice of input signal and different nonlinearities will also be demonstrated via Monte Carlo simulations. The setup is as follows: for each level of SNR, 100 datasets consisting of 300 data points for identification are collected, where a new input and noise sequence is sampled at each instance. The number of regressors and slices used in the experiments are fixed to $n_b = 13$ and $J = \lfloor \sqrt{300} \rfloor = 17$, respectively

Example 1: Elliptically distributed input signal

Now, let the input to the system (16) be given by

$$u(t) = G_u(q)\eta_u(t), \quad \eta_u(t) \sim \mathcal{N}(0, 0.15), \quad (18)$$

where the linear filter has the impulse response

$$g_{u,i} = \begin{cases} 2/3^i, & i = 0, 1, \dots, 5, \\ 0, & \text{otherwise.} \end{cases}$$

Then $\varphi(t)$, defined in (6), is elliptically distributed and the inverse regression methods are expected to work well.

To show the effects of the nonlinearity in (16) three different noninvertible functions are considered. The first nonlinearity is a saturation $f(x) = \text{sat}(x)$ defined as

$$\text{sat}(x) = \begin{cases} x, & \text{if } |x| \leq 1 \\ \text{sign}(x), & \text{otherwise} \end{cases} \quad (19)$$

The identification results of the different methods for this setup is given in Figure 3. The left plot shows the angle (14) between the true impulse response and the estimates for different levels of SNR. The thick gray lines show the mean value over the 100 Monte Carlo simulations while the thin lines show the best and the worst values attained, respectively. The plot to the right shows the resulting impulse response estimates for 20 dB SNR with 3 standard deviation errors when all estimates have been scaled to coincide for the first nonzero coefficient.

For this nonlinearity, we notice that all methods perform quite well with a slight disadvantage for the DR method. The simplicity of the nonlinearity seems to make the improved estimation of the conditional expectation used in LIR, compared to simple estimate used in SIR, unnecessary.

To make things more interesting, we now change the nonlinearity to $f(x) = x \cos(x)$ which is also noninvertible but more oscillatory than the saturation. The results are given in Figure 4. Here, the DR estimate has a significant advantage over

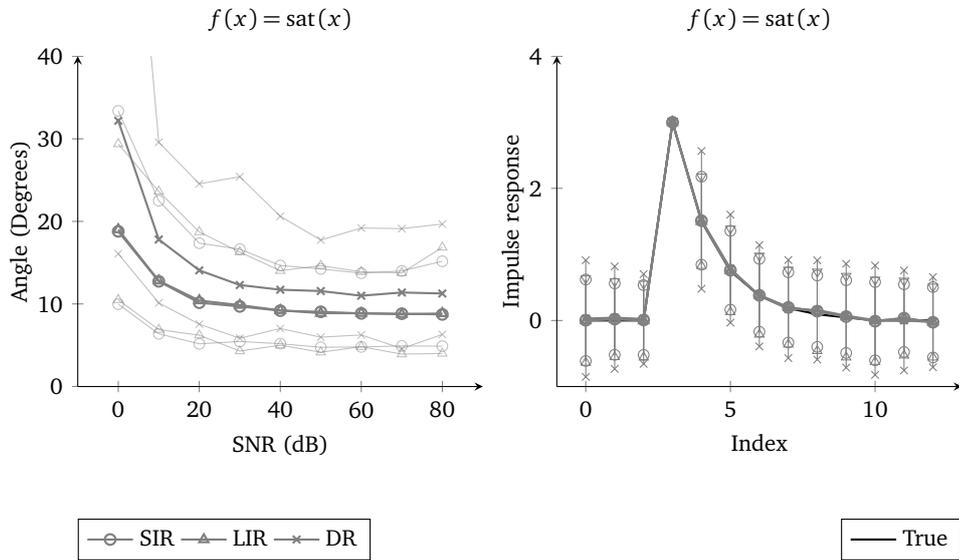


Figure 3: The plot to the left shows the angle (14) between the true impulse response and the estimates for different levels of the SNR (15). The right hand plot shows the mean value of the impulse response estimates for 20 dB SNR with 3 standard deviation errors.

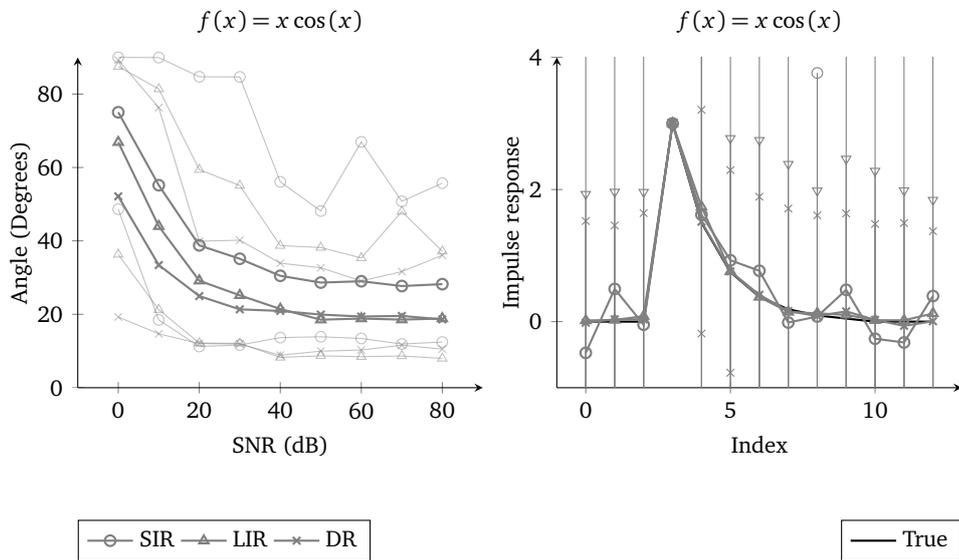


Figure 4: The plot to the left shows the angle (14) between the true impulse response and the estimates for different levels of the SNR (15). The right hand plot shows the mean value of the impulse response estimates for 20 dB SNR with 3 standard deviation errors.

SIR and also over LIR for low SNR. Furthermore, the extra amount of computations required by LIR compared to SIR seems to improve the estimate considerably.

Now, consider the nonlinearity to $f(x) = x^2$ which is symmetric about the origin and $E(\varphi(t)|y(t)) = 0$. Thus, SIR and LIR are expected to yield poor estimates, which is confirmed in Figure 5 where estimates of the impulse responses are given.

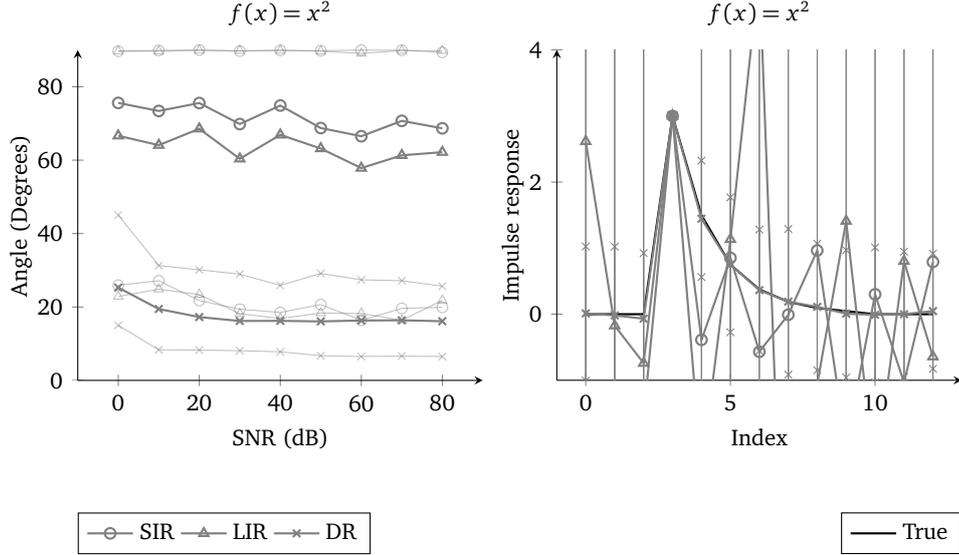


Figure 5: The plot to the left shows the angle (14) between the true impulse response and the estimates for different levels of the SNR (15). The right hand plot shows the mean value of the impulse response estimates for 20 dB SNR with 3 standard deviation errors.

Example 2: Non-elliptically distributed input signal

Now, let the input to the system (16) be given by

$$u(t) = \sin(2\pi\eta_u(t)), \quad \eta_u(t) \sim \mathcal{U}(0, 1), \quad (20)$$

which is non-elliptically distributed and for which there is no guarantee that the inverse regression methods will work. The identification results for the saturation nonlinearity (19) are given in Figure 6.

Here, the methods seem to perform equally well with a slight disadvantage for DR. Furthermore, there seems to be a bias in the estimates of the impulse response which is due to the violation of Assumption 1. Thus, even though the linear design condition is violated the methods presented in this paper might be useful as an initialization method for more advanced methods.

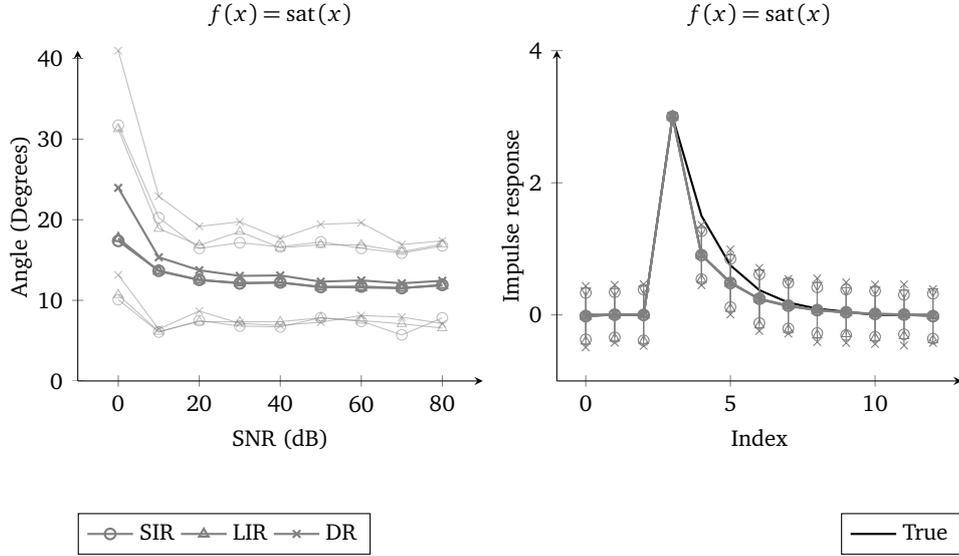


Figure 6: The plot to the left shows the angle (14) between the true impulse response and the estimates for different levels of the SNR (15). The right hand plot shows the mean value of the impulse response estimates for 20 dB SNR with 3 standard deviation errors.

4.2 A multiple branch Wiener system

Example 3

Consider a system in the form

$$y(t) = f(G_1(q)u(t) + w_1(t), G_2(q)u(t) + w_2(t)) + v(t), \quad (21)$$

with the linear subsystems

$$G_1(z) = \frac{1}{z - 0.75}, \quad G_2(z) \triangleq \frac{z^2 - 0.5z}{z^2 - z + 0.5}, \quad (22)$$

and the nonlinearity

$$f(x_1, x_2) = \tanh(x_1) \text{sat}(x_2) + \text{sign}(x_2). \quad (23)$$

The input signal is generated according to

$$u(t) = \frac{1}{1 + 0.6q^{-1}} \eta_u(t), \quad \eta_u(t) \sim \mathcal{N}(0, 1), \quad (24)$$

and the measurement noise $v(t) \sim \mathcal{N}(0, 0.1)$. On each channel, process noise is added according to (21) which is given by

$$w_1(t) = \frac{1}{z + 0.25} \eta_{w,1}(t), \quad \eta_{w,1}(t) \sim \mathcal{N}(0, 0.1), \quad (25)$$

$$w_2(t) = \frac{1}{z - 0.3} \eta_{w,2}(t), \quad \eta_{w,2}(t) \sim \mathcal{N}(0, 0.1). \quad (26)$$

The simulations are as follows, for 100 datasets consisting of 3,000 data points for identification are collected with new realizations of both input and noise signals at each instance. The resulting angles (14), for the choice $d = 2$, between the truncated true impulse response and the estimates are given in Table 1.

Table 1: The angles (14) between the truncated true impulse response and the estimates for Example 3.

	Mean	Min	Max
SIR	3.91	2.35	7.05
LIR	3.85	2.37	6.45
DR	4.46	2.34	7.88

Here, all methods have similar performance with a slight advantage for LIR. The choice $d = 2$ can be motivated by observing the plot of the singular values that is calculated in Step 3 of Algorithm 1. These are presented in Figure 7 where there is a gap between the singular values from index 2 to 3 for all the methods.

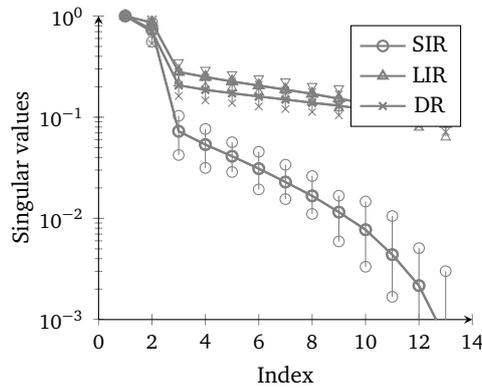


Figure 7: The singular values for Example 3 with the thick line representing the mean value and the bars indicating the maximum and minimum values attained respectively.

5 Concluding remarks

In this paper, methods from inverse regression has evaluated for the identification of systems from the Wiener class. The results on artificial data looks promising and the methods might be used to initialize more advanced methods.

One method that seems to perform well for most nonlinearities is the DR method. This method is not fully understood in the case of dynamically dependent data and needs to be studied further in the future.

Recently, the ideas of inverse regression has been generalized to reproducing kernel Hilbert spaces and have shown promising results for different problems in machine learning, see, for instance, Wu [2008] for a generalization of SIR and Kim and Pavlovic [2011] for a method that does not rely on discretizing the output.

These methods could improve the estimation results, especially when the input signal has a more complex distribution, and is a topic for future research.

References

- S. Boyd and L. O. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on Circuits and Systems*, 32(11), 1985.
- E. Bura. Using linear smoothers to assess the structural dimension of regressions. *Statistica Sinica*, 13, 2003.
- E. Bura and R. D. Cook. Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society: Series B*, 63(2):393–410, 2001.
- E. Bura and J. Yang. Dimension estimation in sufficient dimension reduction: A unifying approach. *Journal of Multivariate Analysis*, 102(1):130–142, 2011.
- J. J. Bussgang. Crosscorrelation functions of amplitude-distorted Gaussian signals. Technical Report 216, MIT Research Laboratory of Electronics, Cambridge, Massachusetts, 1952.
- R. D. Cook and S. Weisberg. Comments on "Sliced inverse regression for dimension reduction". *Journal of the American Statistical Association*, 86(414):328–332, 1991.
- R. D. Cook, L. M. Forzani, and D. R. Tomassi. LDR: A package for likelihood-based sufficient dimension reduction. *Journal of Statistical Software*, 39(3), 2011.
- M. L. Eaton. A characterization of spherical distributions. *Journal of Multivariate Analysis*, 20(2):272–276, 1986.
- J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.
- J. H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82(397), 1987.
- J. H. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American Statistical Association*, 76(376):817–823, 1981.
- F. Giri and E. Bai, editors. *Block-Oriented Nonlinear System Identification*. Lecture Notes in Control and Information Sciences. Springer Verlag, 2010.
- M. Kim and V. Pavlovic. Central subspace dimensionality reduction using covariance operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):657–670, 2011.
- B. Li and S. Wang. On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102(479):997–1008, 2007.
- B. Li, H. Zha, and F. Chiaromonte. Contour regression: a general approach to dimension reduction. *The Annals of Statistics*, 33(4):1580–1616, 2005.

- K. C. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.
- K. C. Li. On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. *Journal of the American Statistical Association*, 87(420):1025–1029, 1992.
- L. Ljung. *System Identification: Theory for the User*. Prentice Hall, second edition, 1999.
- C. Lyzell and M. Enqvist. Sliced inverse regression for the identification of dynamical systems. Technical Report LiTH-ISY-R-3031, Department of Electrical Engineering, Linköping University, 2011.
- A. H. Nuttall. Theory and application of the separable class of random processes. Technical Report 343, MIT Research Laboratory of Electronics, Cambridge, Massachusetts, 1958.
- R. K. Pearson. *Discrete-Time Dynamical Models*. Oxford University Press, 1999.
- M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley-Interscience, 1980.
- M. Schetzen. Nonlinear system modelling and analysis from the Volterra and Wiener perspective. In F. Giri and E. Bai, editors, *Block-Oriented Nonlinear System Identification*, Lecture Notes in Control and Information Sciences. Springer Verlag, 2010.
- H. M. Wu. Kernel sliced inverse regression with applications on classification. *Journal of Computational and Graphical Statistics*, 17(3):590–610, 2008.

	Avdelning, Institution Division, Department	Datum Date
	Division of Automatic Control Department of Electrical Engineering	2011-11-14
Språk Language <input type="checkbox"/> Svenska/Swedish <input checked="" type="checkbox"/> Engelska/English <input type="checkbox"/> _____	Rapporttyp Report category <input type="checkbox"/> Licentiatavhandling <input type="checkbox"/> Examensarbete <input type="checkbox"/> C-uppsats <input type="checkbox"/> D-uppsats <input checked="" type="checkbox"/> Övrig rapport <input type="checkbox"/> _____	ISBN _____ ISRN _____ Serietitel och serienummer ISSN Title of series, numbering <u>1400-3902</u>
URL för elektronisk version http://www.control.isy.liu.se		LiTH-ISY-R-3032
Titel Title	Inverse Regression for the Wiener Class of Systems	
Författare Author	Christian Lyzell, Martin Enqvist	
Sammanfattning Abstract		
<p>The concept of inverse regression has turned out to be quite useful for dimension reduction in regression analysis problems. Using methods like <i>sliced inverse regression</i> (SIR) and <i>directional regression</i> (DR), some high-dimensional nonlinear regression problems can be turned into more tractable low-dimensional problems. Here, the usefulness of inverse regression for identification of nonlinear dynamical systems will be discussed. In particular, it will be shown that the inverse regression methods can be used for identification of systems of the Wiener class, that is, systems consisting of a number of parallel linear subsystems followed by a static multiple-input single-output nonlinearity. For a particular class of input signals, including Gaussian signals, the inverse regression approach makes it possible to estimate the linear subsystems without knowing or estimating the nonlinearity.</p>		
Nyckelord Keywords System identification; Dimension reduction; Inverse regression.		