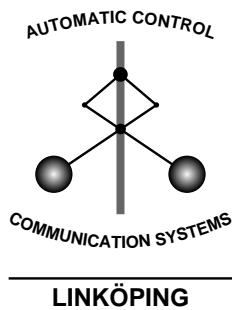


# A Non-Asymptotic Approach to Local Modelling

Jacob Roll, Alexander Nazin, Lennart Ljung

Division of Automatic Control  
Department of Electrical Engineering  
Linköpings universitet, SE-581 83 Linköping, Sweden  
WWW: <http://www.control.isy.liu.se>  
E-mail: [roll,ljung@isy.liu.se](mailto:roll,ljung@isy.liu.se), [nazine@ipu.rssi.ru](mailto:nazine@ipu.rssi.ru)

January 9, 2003



Report no.: LiTH-ISY-R-2482

Submitted to CDC'02

Technical reports from the Control & Communication group in Linköping are available at <http://www.control.isy.liu.se/publications>.



# A Non-Asymptotic Approach to Local Modelling

Jacob Roll\*, Alexander Nazin†, Lennart Ljung\*

January 9, 2003

## Abstract

Local models and methods construct function estimates or predictions from observations in a local neighborhood of the point of interest. The bandwidth, i.e., how large the local neighborhood should be, is often determined based on asymptotic analysis. In this paper, an alternative, non-asymptotic approach that minimizes a uniform upper bound on the mean square error for a linear estimate is proposed. It is shown, for the scalar case, that the solution is obtained from a quadratic program, and that it maintains many of the key features of the asymptotic approaches. Moreover, examples show that the proposed approach in some cases is superior to an asymptotically based local linear estimator.

## 1 Introduction: Local Models

Non-linear black box models of dynamical systems have long been of central interest in system identification, see, e.g. the survey [1]. In the control community mostly models of function expansion type have been applied, like Artificial Neural Network (ANN) models, wavelets, and (neuro-)fuzzy models. In statistics, the interest has been focused on various local methods, like kernel methods, [2], local polynomial approaches, [3] and trees, [4].

A *local* model or method builds the function estimate or prediction from observations in a local neighborhood of the point of interest. Also most function expansion methods are of this character: A radial basis neural network is built up from basis functions with local support, and the standard sigmoidal (one hidden layer feed-forward) network is local around certain hyperplanes in the regressor space.

No matter what type of local modelling approach is taken, the central problem is the *bandwidth* question:

- How big should the local neighborhood be?

This is a variant of the classical bias/variance trade-off. Intuitively, it is clear that the answer must depend on three items:

1. How many data are available (and how are they distributed)?

---

\*Div. of Automatic Control, Linköping University, SE-58183 Linköping, Sweden, e-mail: roll, ljung@isy.liu.se

†Institute of Control Sciences, Profsoyuznaya str., 65, 117997 Moscow, Russia, e-mail: nazine@ipu.rssi.ru

2. How smooth is the function surface (supposed to be)?
3. How much noise is there in the observations?

This problem has been studied extensively in the statistical literature, and there are several solutions based on asymptotic (in the number of observations) analysis.

In this paper we propose another solution that is not based on the asymptotic behavior of the estimates. Based on a smoothness measure and noise variance, we compute a uniform upper bound of the mean square error (MSE) of a linear estimate, as a function of the estimator parameters. This upper bound is then minimized. It turns out that this problem can be reformulated as a quadratic programming (QP) problem, which can be solved efficiently. It also turns out that this solution has many of the key features of the asymptotically optimal estimators, but for finite number of observations it produces better guaranteed error bounds.

The paper is organized as follows: In Section 2 we give a brief overview of some of the existing local approaches. The new approach is described in Section 3, and some of its most important properties are derived in Section 4. In Section 5 it is shown that the improvement in model quality could be substantial with the new approach. Some extensions are described in Section 6 and conclusions are given in Section 7.

## 2 Kernel Methods and the Local Polynomial Approach

Let us consider the problem of estimating the value  $f(x_0)$  of an unknown scalar function  $f : \mathbf{R} \rightarrow \mathbf{R}$  at a given point  $x_0$ , given a set of input-output pairs  $\{(x_k, y_k)\}_{k=1}^N$ , coming from the relation

$$y_k = f(x_k) + e_k \quad (1)$$

We assume that the function  $f$  is continuously differentiable, and that there is a Lipschitz constant  $L$  such that

$$|f'(x_1) - f'(x_2)| \leq L|x_1 - x_2| \quad (2)$$

Denote the class of functions satisfying these assumptions by  $\mathcal{F}_L$ .

The noise terms  $e_k$  are independent random variables with  $Ee_k = 0$  and  $Ee_k^2 = \sigma_k^2$ . Both  $L$  and  $\sigma_k$  are assumed to be positive constants, given a priori. We also introduce

$$\tilde{x}_k = x_k - x_0 \quad (3)$$

and  $X = (x_1, \dots, x_N)$ .

A common approach for this estimation problem is to use a linear estimator on the form

$$\hat{f}_{x_0} = \sum_{k=1}^N \alpha_k y_k \quad (4)$$

where  $\hat{f}_{x_0}$  is our estimate of  $f(x_0)$ . The problem then reduces to finding good weights  $\alpha_k$ , which give reasonably small bias and variance of the estimate over

the class  $\mathcal{F}_L$ . This will be done by considering the *linear minimax risk* (see [3]), defined by

$$R(X, L) = \inf_{\hat{f}_{x_0}} \sup_{f \in \mathcal{F}_L} E[(\hat{f}_{x_0} - f(x_0))^2 | X] \quad (5)$$

where the infimum is taken over all linear estimators.

A classic family of methods to decide the weights are the *kernel methods*. Here, a *kernel function*  $K$ , which usually is a symmetric probability density function, is used to determine the weights. The width of the kernel is determined by introducing a bandwidth parameter  $h$ , and letting  $K_h(\cdot) = K(\cdot/h)/h$ . As an example of a kernel method, the *Nadaraya-Watson estimator* [2, 5] is given by

$$\hat{f}_{x_0}^{NW} = \frac{\sum_{k=1}^N K_h(\tilde{x}_k) y_k}{\sum_{k=1}^N K_h(\tilde{x}_k)} \quad (6)$$

The choice of bandwidth is a bias/variance trade-off problem. For more details about this, asymptotic properties, etc., see [3].

Some common choices of kernel functions are the Gaussian kernel and the *Epanechnikov kernel*

$$K(u) = \frac{3}{4}(1 - u^2)_+ \quad (7)$$

where  $(\cdot)_+ = \max\{\cdot, 0\}$ .

An alternative to the described method is the *local polynomial modelling* approach. In this approach, the estimator is determined by locally fitting a polynomial to the given data via minimization of the weighted least-squares problem:

$$\hat{\beta} = \arg \min_{\beta} \sum_{k=1}^N K_h(\tilde{x}_k) \left( y_k - \sum_{j=0}^p \beta_j \tilde{x}_k^j \right)^2 \quad (8)$$

The resulting estimator is obtained as  $\hat{f}_{x_0}^{LP} = \hat{\beta}_0$ . When  $p = 0$ ,  $\hat{f}_{x_0}^{LP}$  will be the Nadaraya-Watson estimator. When  $p = 1$ , the estimator is called a *local linear estimator*, and can be expressed explicitly as

$$\hat{f}_{x_0}^{LP} = \frac{\sum_{k=1}^N w_k y_k}{\sum_{k=1}^N w_k}, \quad (9)$$

$$w_k = K_h(\tilde{x}_k) \left( \sum_{i=1}^N K_h(\tilde{x}_i) \tilde{x}_i^2 - \tilde{x}_k \sum_{i=1}^N K_h(\tilde{x}_i) \tilde{x}_i \right)$$

One can show that (see [3]), under mild assumptions and as  $N \rightarrow \infty$ , the local linear estimator asymptotically achieves the linear minimax risk, if the Epanechnikov kernel is used with a bandwidth given by

$$h = \left( \frac{15\sigma^2}{p_x(x_0)L^2N} \right)^{1/5} \quad (10)$$

where  $p_x$  is the probability density of the observations  $x_k$  and  $\sigma_k = \sigma$  for all  $k = 1, \dots, N$ .

The different methods given in this section for choosing the weights  $\alpha_k$  in the linear estimator (4) were all justified using asymptotic arguments, as  $N \rightarrow \infty$ .

However, in reality only a finite number of data is given. Furthermore, these data may be sparsely and non-uniformly distributed. This might deteriorate the performance of the estimation methods.

In the following section, we will present a non-asymptotic approach for determining the weights, based on a uniform (over  $\mathcal{F}_L$ ) upper bound on the MSE (and hence on the linear minimax risk).

### 3 A Non-Asymptotic, Min-Max Approach

We will again use a linear estimator on the form (4). First, let us pose some requirements on the weights  $\alpha_k$ . We require

$$\sum_{k=1}^N \alpha_k = 1 \quad (11a)$$

$$\sum_{k=1}^N \alpha_k \tilde{x}_k = 0 \quad (11b)$$

Under these restrictions, any linear function is estimated with zero bias. Moreover, if (11) does not hold, the bias is unbounded over  $\mathcal{F}_L$  (see [6]). Note also that the local linear estimator described in the last section will automatically satisfy these requirements.

From (2), we get

$$f(x) = f(x_0) + f'(x_0)\tilde{x} + \frac{c(x)}{2}\tilde{x}^2 \quad (12)$$

for some  $c(x)$ ,  $|c(x)| \leq L$ . Using all these equations, we now get

$$\begin{aligned} \hat{f}_{x_0} &= \sum_{k=1}^N \alpha_k \left( f(x_0) + f'(x_0)\tilde{x}_k + \frac{c(x_k)}{2}\tilde{x}_k^2 + e_k \right) = \\ &= f(x_0) + \sum_{k=1}^N \alpha_k \frac{c(x_k)}{2}\tilde{x}_k^2 + \sum_{k=1}^N \alpha_k e_k \end{aligned}$$

and an upper bound on the MSE (over  $\mathcal{F}_L$ )

$$\begin{aligned} E[(\hat{f}_{x_0} - f(x_0))^2 | X] &= \left( \sum_{k=1}^N \alpha_k \frac{c(x_k)}{2}\tilde{x}_k^2 \right)^2 + \sum_{k=1}^N \alpha_k^2 \sigma_k^2 \\ &\leq \frac{L^2}{4} \left( \sum_{k=1}^N \tilde{x}_k^2 |\alpha_k| \right)^2 + \sum_{k=1}^N \alpha_k^2 \sigma_k^2 \end{aligned} \quad (13)$$

Note that this bound is tight and attained by a parabola with  $f''(x) = L$  if the weights  $\alpha_k$  are non-negative.

It is now natural to minimize the upper bound in (13). Hence, we would like to choose the values of  $\alpha_k$  that minimize the following convex optimization

problem:

$$\begin{aligned}
& \min_{\alpha_k} \frac{L^2}{4} \left( \sum_{k=1}^N \tilde{x}_k^2 |\alpha_k| \right)^2 + \sum_{k=1}^N \alpha_k^2 \sigma_k^2 \\
& \text{subj. to} \quad \sum_{k=1}^N \alpha_k = 1 \\
& \quad \quad \quad \sum_{k=1}^N \alpha_k \tilde{x}_k = 0
\end{aligned} \tag{14}$$

This problem is equivalent to the following QP:

$$\begin{aligned}
& \min_{\alpha_k, s_k} \frac{L^2}{4} \left( \sum_{k=1}^N \tilde{x}_k^2 s_k \right)^2 + \sum_{k=1}^N s_k^2 \sigma_k^2 \\
& \text{subj. to} \quad s_k \geq \alpha_k \\
& \quad \quad \quad s_k \geq -\alpha_k \\
& \quad \quad \quad \sum_{k=1}^N \alpha_k = 1 \\
& \quad \quad \quad \sum_{k=1}^N \alpha_k \tilde{x}_k = 0
\end{aligned} \tag{15}$$

The equivalence between (14) and (15) can be shown as follows: Given a feasible solution  $\alpha = (\alpha_1, \dots, \alpha_N)$  to (14), we can get a feasible solution to (15) with the same value of the objective function by using the same  $\alpha$  and  $s_k = |\alpha_k|$ . Hence (15) is a relaxation of (14), and it suffices to show that when the optimal value of (15) is reached, then  $s_k = |\alpha_k|$  for all  $k = 1, \dots, N$ . Suppose, e.g., that  $s_1 > |\alpha_1|$ . Then, without changing any other variables, the value of the objective function can be reduced by decreasing  $s_1$ . This can be seen by observing that the coefficient before  $s_1$  in the first sum of the objective function is non-negative, and the coefficient before  $s_1^2$  in the second sum is positive, so decreasing  $s_1$  will decrease the objective function. Therefore, when the objective function will reach its minimum, then  $s_k = |\alpha_k|$ , and the equivalence is shown.

It should be pointed out, that the fact that the upper bound in (13) is tight for non-negative weights  $\alpha_k$  does not necessarily mean that minimizing (15) yields the weights that minimize the worst-case MSE, even if the resulting weights are positive. The reason for this is that the weights that really minimize the worst-case MSE may be negative, and so the upper bound is not tight for these weights. However, preliminary experimental results show that the solution of (15) gives an upper bound which is mostly within a few percents from the optimal upper bound.

Solving the QP (15) can be done very efficiently using standard solvers, e.g., CPLEX [7].

## 4 Some Basic Properties of the Approach

Since the QP minimax approach minimizes an upper bound on the linear minimax risk, one would expect that the weights  $\alpha_k$  would asymptotically converge

to the weights of the local linear estimator using the Epanechnikov kernel with a bandwidth given by (10). This is also the case under certain assumptions, as the following theorem shows. For a proof of the result, see [6].

**Theorem 4.1** *Consider the problem of estimating an unknown function  $f \in \mathcal{F}_L$  at a given internal point  $x_0 \in (-1/2, 1/2)$  under a fixed, equidistant design model as follows*

$$x_k = \frac{k-1}{N-1} - \frac{1}{2}, \quad k = 1, \dots, N \quad (16)$$

and with  $\sigma_k = \sigma > 0$  for all  $k = 1, \dots, N$ . Let

$$\alpha^{opt} = (\alpha_1^{opt}, \dots, \alpha_N^{opt})$$

be the minimum point of the upper bound function

$$U(\alpha) = \left( \frac{L}{2} \sum_{k=1}^N |\alpha_k| \tilde{x}_k^2 \right)^2 + \sigma^2 \sum_{k=1}^N \alpha_k^2 \quad (17)$$

subject to the constraints (11). Then asymptotically, as  $N \rightarrow \infty$ ,

$$\alpha_k^{opt} \approx \frac{3}{4} C_N \left( 1 - \left( \frac{\tilde{x}_k}{h_N} \right)^2 \right)_+, \quad k = 1, \dots, N \quad (18)$$

where

$$C_N \asymp \frac{1}{Nh_N}, \quad h_N \asymp \left( \frac{15\sigma^2}{L^2N} \right)^{1/5} \quad \text{as } N \rightarrow \infty \quad (19)$$

Hence, the optimal weights (18) approximately coincide with related asymptotically optimal weights and bandwidth of the local polynomial estimator for the worst case function in  $\mathcal{F}_L$ .

Here  $a_N \asymp b_N$  means asymptotic equivalence of two real sequences  $(a_N)$  and  $(b_N)$ , that is  $a_N/b_N \rightarrow 1$  as  $N \rightarrow \infty$ .

**Remark.** When the data are symmetrically distributed around  $x_0$ , e.g., when  $x_0 = 0$ , the relation (18) will hold exactly also for finite  $N$ . In other words, the non-zero weights will lie along a parabola given by (18), with an asymptotic bandwidth given by (19).

Another interesting feature of the QP minimax approach is that in most cases, the weights  $\alpha_k$  corresponding to  $x_k$  lying beyond a certain distance from  $x_0$  will be zero (see, e.g., the example in Section 5). This can be regarded as an *automatic bandwidth selection*, which means that the user does not have to bother about how many of the samples should be included in the estimator. In fact, the following theorem holds:

**Theorem 4.2** *Suppose that the problem (15) is feasible, and  $\sigma_k = \sigma > 0$  for all  $k = 1, \dots, N$ . Then there exist three numbers  $\mu_1 > 0$ ,  $\mu_2$ , and  $\mu_3 \geq 0$ , such that for an optimal solution  $(s^{opt}, \alpha^{opt})$ , we have*

$$\alpha_k^{opt} = \begin{cases} \mu_1 + \mu_2 \tilde{x}_k - \mu_3 \tilde{x}_k^2, & \mu_3 \tilde{x}_k^2 \leq \mu_1 + \mu_2 \tilde{x}_k \\ 0, & |\mu_1 + \mu_2 \tilde{x}_k| \leq \mu_3 \tilde{x}_k^2 \\ \mu_1 + \mu_2 \tilde{x}_k + \mu_3 \tilde{x}_k^2, & \mu_1 + \mu_2 \tilde{x}_k \leq -\mu_3 \tilde{x}_k^2 \end{cases} \quad (20)$$

Moreover, if there are two indices  $k_1$  and  $k_2$ , such that  $0 \neq \tilde{x}_{k_1} \neq \tilde{x}_{k_2} \neq 0$ , then  $\mu_3 > 0$ .



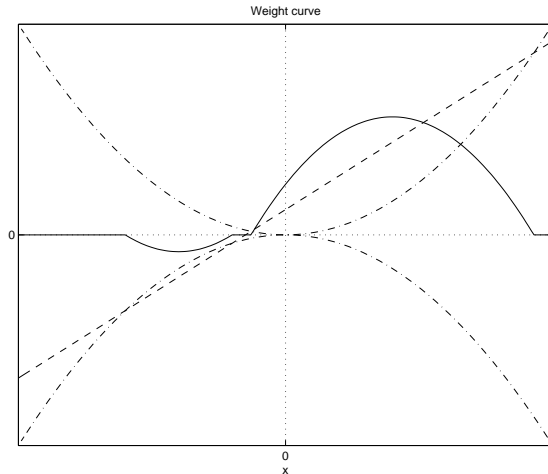


Figure 1: Principal shape of the weight curve (solid curve). The dash-dotted parabolas are  $\pm\mu_3\tilde{x}^2$ , and the dashed line is  $\mu_1 + \mu_2\tilde{x}$ . (The weight curve is scaled by a factor 4 to make the figure more clear.)

*Proof:* See [6].

Figure 1 shows the principal shape of the curve along which the weights  $\alpha_k^{opt}$  are placed. When  $\mu_3\tilde{x}_k^2 \leq \mu_1 + \mu_2\tilde{x}_k$  (which in the figure corresponds to the dashed line being above the upper dash-dotted parabola), the weights will be positive. When  $\mu_1 + \mu_2\tilde{x}_k \leq -\mu_3\tilde{x}_k^2$  (the dashed line is below the lower dash-dotted parabola), the weights are negative, and otherwise they are zero.

The last result opens up for a possible reduction of the computational complexity: Since many of the weights  $\alpha_k$  will be zero, we can already on beforehand exclude data that will most likely correspond to zero weights, thus making the QP (15) considerably smaller. Having solved (15), one can easily check whether or not the excluded weights really should be zero, by checking if the excluded data points satisfy  $|\mu_1 + \mu_2\tilde{x}_k| \leq \mu_3\tilde{x}_k^2$  (the middle case of (20)). For more details, see [6].

## 5 An Example

To illustrate a possible situation where the QP minimax approach described in Section 3 might be superior to the local linear estimator, we consider the following example. Let

$$f(x) = x^2 \sin(2x) \quad (21)$$

and let  $x_k, k = 1, \dots, 50$ , be taken from a uniform distribution on  $[-2, 2]$ . We let  $\sigma_k^2 = 1$  and  $L = 13$ . Suppose that we would like to estimate  $f(0)$ . One example is given in Figure 2. The prediction errors, worst-case bias, variance, and MSE for this particular set of data  $\{(x_k, y_k)\}_{k=1}^{50}$  are listed in Table 1 for the two different approaches. Note that all values, except for the prediction errors, only depend on the distribution of  $x_k$ , not on  $y_k$ .

As we can see in Figure 2(b), most weights are zero, both in the local linear and QP approach. In the local linear estimator, the kernel function together

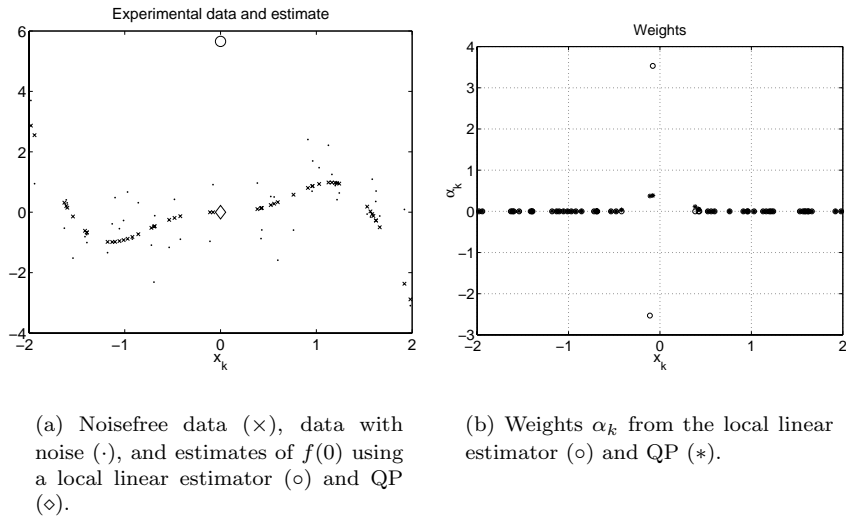


Figure 2: Comparison of the local linear estimator, using the Epanechnikov kernel with bandwidth given by (10), and the QP minimax approach.

Table 1: Comparison of the asymptotically optimal local linear estimator and the QP minimax approach for the example shown in Figure 2.

	Epanechnikov	QP
Prediction error	-5.6567	-0.0016
(Bias) <sup>2</sup> (upper bound)	0.1148	0.0942
Variance	18.8892	0.3044
MSE (upper bound)	19.0040	0.3987

with the bandwidth decides which weights should be zero. In the QP minimax approach, however, this is taken care of by the automatic bandwidth selection mentioned in Section 4, which allows for a greater flexibility, such that the minimal upper bound on the MSE can be achieved. In this specific example, the effect is that the local linear estimator just takes the two data points closest to the point of interest (which both happen to be negative) into account. This makes the estimate very sensitive to the actual noise realization. The QP minimax approach, on the other hand, takes three additional points into account, of which two are positive, thereby making the estimate much less noise sensitive.

Table 2 shows the resulting estimates of the actual MSE from four Monte-Carlo simulations (with 10000 experiments each), where  $x_k$  are taken from a uniform distribution on  $[-2, 2]$  and  $\sigma_k^2 = 1$ .  $f(x_0)$  is estimated for  $x_0 = 0$  and  $x_0 = 1.5$ , using  $N = 20$  or  $N = 50$  observations. As can be seen, the QP minimax approach performs better than the local linear approach in all four cases, and the difference is accentuated when the number of data is small, just as expected.

Table 2: Comparison of the asymptotically optimal local linear estimator and the QP minimax approach; results from Monte-Carlo simulations.

$x_0$	N	MSE (Epanechnikov)	MSE (QP)
0	50	0.2240	0.1424
1.5	50	0.2289	0.1785
0	20	26.9244	0.3182
1.5	20	9.5489	0.5087

## 6 Extensions

### 6.1 Prior knowledge

One may think of the constraints (11a) and (11b) as *ad hoc*, although reasonable. We can remove these constraints and instead consider a restricted family of functions with some prior knowledge of the function value and its derivative:

$$|f(x_0) - a| \leq A, \quad |f'(x_0) - b| \leq B \quad (22)$$

and formulate the corresponding min-max MSE problem. It can then be shown, see [6], that for sufficiently large  $A$  and  $B$ , the solution will automatically be subject to the constraints (11a) and (11b).

Moreover, if we form a reasonable linear estimate of the derivative  $f'(x_0)$  and plug that in as a known value (with or without error), the resulting estimator will obey (11b).

### 6.2 More smoothness

The criterion we have used assumes the function to be once continuously differentiable with a bound on the Lipschitz constant of the derivative. It is quite easy to extend this to any degree of assumed differentiability. The min-max problem will still be reducible to a QP problem, [6].

### 6.3 Higher regressor dimensions

In this paper we have assumed that the function to be estimated has a scalar argument. In most applications, in particular to dynamic systems, the regressors will have a higher dimension. The extension to this case is immediate. The QP problem is still over scalar weights and does not become more complex. See, again, [6].

## 7 Conclusions

The proposed non-asymptotic min-max approach to local modelling has a number of interesting features:

- The problem is phrased without any reference to bandwidth. The formulation offers the possibility to use all observations. The solution to the QP problem however shows that there is a bandwidth feature even for a finite number of measurements: Observations outside a certain band carry weights that are exactly zero.

- Although our approach does not give strictly better estimates (in the MSE sense) than, say, the local polynomial approach in all cases, the important point is that the delivered guaranteed MSE bound is strictly better than what other approaches can offer. In practice it is of course only this guaranteed bound that can be used for confidence intervals etc, since the actual MSE depends on the unknown function.
- The improvement over asymptotically optimal estimates is more pronounced (naturally enough) for fewer data, and more non-uniformly distributed observation points  $x_k$ . For applications to higher regressor dimensions and dynamical systems, this a very valuable property.

Finally, one may ask how to find the “input values”  $L$  and  $\sigma_k^2$  to the algorithms. This is the same problem as for the kernel methods and the local polynomial approach, and has been extensively studied in the statistical literature. We refer to [3] for ideas around this.

The application of these methods to dynamical systems with the regressors  $x_k$  being built up by past inputs and outputs is straightforward. The method can be used as an alternative to building non-linear black-box models in a “Model-On-Demand” fashion and applied to, for example, model predictive control. See [8] for such ideas.

## References

- [1] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Y. Glorennec, H. Hjalmarsson, and A. Juditsky, “Nonlinear black-box modeling in system identification: a unified overview,” *Automatica*, vol. 31, no. 12, pp. 1691–1724, 1995.
- [2] E. Nadaraya, “On estimating regression,” *Theory of Probability and its Applications*, vol. 10, pp. 186–190, 1964.
- [3] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman & Hall, 1996.
- [4] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth & Brooks, 1984.
- [5] G. Watson, “Smooth regression analysis,” *Sankhyā*, vol. A, no. 26, pp. 359–372, 1964.
- [6] J. Roll, A. Nazin, and L. Ljung, “Non-asymptotic optimal guaranteed MSE function estimation,” Tech. Rep. LiTH-ISY-R-2415, Department of Electrical Engineering, Linköping University, 2002.
- [7] ILOG, Inc., *CPLEX 7.0 User’s Manual*, Gentilly, France, 2000.
- [8] A. Stenman, *Model on Demand: Algorithms, Analysis and Applications*, Ph.D. thesis, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, 1999.