LOCAL MODELLING WITH A PRIORI KNOWN BOUNDS USING DIRECT WEIGHT OPTIMIZATION

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

* 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`

Keywords: Local modelling, non-parametric identification, nonlinear systems, minimax techniques, convex programming.

Abstract

In local modelling, function estimates are computed from observations in a local neighborhood of the point of interest. A central question is how to choose the size of the neighborhood. Often this question has been tackled using asymptotic (in the number of observations) arguments. The recently introduced direct weight optimization approach is a non-asymptotic approach, minimizing an upper bound on the mean squared error. In this paper the approach is extended to also take a priori known bounds on the function and its derivative into account. It is shown that the result will sometimes, but not always, be improved by this information. The proposed approach can be applied, e.g., to prediction of nonlinear dynamic systems and model predictive control.

1 Introduction

Local modelling of different types have been of great interest for a long time in system identification and statistics. A local model or method for function approximation or prediction computes the estimate using information from observations in a local neighborhood of the point of interest. Many of the different nonlinear black-box methods are of this type, such as radial basis neural networks, multiple-model approaches etc. (see, e.g., [5, 11]). In statistics, local methods such as kernel methods [6, 14], local polynomial modelling [3], and trees [2] have been popular.

A central question in local modelling is the *bandwidth* question: how to select the size of the neighborhood. This becomes a bias/variance trade-off, which has been studied extensively in the statistics literature, and many solutions based on asymptotic (in the number of observations) arguments have been proposed.

In this paper, a *direct weight optimization* (DWO) approach is considered. This approach is a non-asymptotic approach, where the weights of a linear (or affine) estimator are determined by minimizing a uniform upper bound on the mean squared error (MSE) over function classes having a Lipschitz bound L on the derivative. It turns out that the optimization problem can be formulated as a quadratic program (QP) or a second-order cone program (SOCP). The basic idea was presented in [8, 9] (see also [10] for an early contribution in the same direction), where it was also shown that the approach has several interesting properties:

- Outside a certain region in the regressor space, the weights will automatically become zero. Hence, we automatically get a finite bandwidth.
- Asymptotically (as $N \to \infty$, N being the number of observations), the weights will (under certain assumptions) converge to the weights obtained by local linear modelling with an asymptotically optimal kernel function (see [3]).
- When the Lipschitz constant $L = 0$, i.e., when the true system function is affine, the function estimates obtained are the same as for a global "affine ARX" model.

Here, the approach is extended to also take a priori known bounds on the function and its derivative into account. In practice, such bounds may be given by physical constraints, or by practical experience and expert knowledge. It turns out that the extension is very natural, and some theorems can be given about when we actually benefit from the extra information.

In Section 2, the basic problem is presented for the univariate case, and in Sections 3 and 4 the DWO approach is outlined. Section 5 deals with multivariate functions. An example is given in Section 6, and conclusions in Section 7.

2 Basic problem

Let us consider the problem of estimating the value $f(\varphi_0)$ of an unknown function $f : \mathbf{R} \to \mathbf{R}$ at a given point φ_0, given a set of input-output pairs $\{(\varphi_k, y_k)\}_{k=1}^N$, coming from the relation

$$y_k = f(\varphi_k) + e_k \tag{1}$$

Assume that the function f is continuously differentiable, and that there are known positive constants L, δ, Δ, and known constants a, b such that

$$|f'(\varphi_1) - f'(\varphi_2)| \le L|\varphi_1 - \varphi_2| \quad \forall\, \varphi_1, \varphi_2 \in \mathbf{R} \tag{2}$$

$$|f(\varphi_0) - a| \le \delta\,, \tag{3}$$

$$|f'(\varphi_0) - b| \le \Delta\,. \tag{4}$$

where f' is the derivative of f. Denote the class of functions satisfying these assumptions by $\mathcal{F}_2(L, \delta, \Delta)$[1].

[1] The function class also depends on a and b, but for notational convenience, this dependency is not written out explicitly.

The noise terms e_k are independent random variables with $Ee_k = 0$ and $Ee_k^2 = \sigma_k^2$ where σ_k are assumed to be positive constants, given a priori. For simplicity, only constant variance (i.e., $\sigma_k^2 = \sigma^2$) will be considered in this paper. We also assume that e_k and φ_j are independent for all j, k. The notation

$$\tilde{\varphi}_k = \varphi_k - \varphi_0 \tag{5}$$

and $X = (\varphi_1, \ldots, \varphi_N)$ will also be used.

There are some particular cases that deserve special attention:

- If $\delta \to +\infty$ then the limit class

$$\mathcal{F}_2(L, \Delta) \triangleq \mathcal{F}_2(L, \infty, \Delta) = \bigcup_{t=1}^{\infty} \mathcal{F}_2(L, \delta, \Delta)|_{\delta=t} \tag{6}$$

  describes the situation where we have no direct a priori information on function value $f(\varphi_0)$.

- If both $\delta \to +\infty$ and $\Delta \to +\infty$ then the limit class

$$\mathcal{F}_2(L) \triangleq \mathcal{F}_2(L, \infty) = \bigcup_{t=1}^{\infty} \mathcal{F}_2(L, \Delta)|_{\Delta=t} \tag{7}$$

  represents a set of continuously differentiable functions meeting the only condition (2). This case was studied previously in [8, 9].

A common approach for the given estimation problem is to use a *linear estimator*

$$\hat{f}(\varphi_0) = \sum_{k=1}^{N} w_k y_k \tag{8}$$

The problem then reduces to finding good weights w_k, according to some criterion. In the following, we will consider the slightly more general class of *affine estimators*:

$$\hat{f}(\varphi_0) = w_0 + \sum_{k=1}^{N} w_k y_k \tag{9}$$

The performance of an estimator $\hat{f}(\varphi_0)$ will be evaluated by the *worst-case mean squared error (MSE)* defined by

$$V(\mathcal{F}, w_0, w) = \sup_{f \in \mathcal{F}} MSE(f, w_0, w) \tag{10}$$

where $w = \begin{pmatrix} w_1 & \ldots & w_N \end{pmatrix}^T$ and

$$MSE(f, w_0, w) = E[(\hat{f}(\varphi_0) - f(\varphi_0))^2 | X] \tag{11}$$

In Section 3, upper bounds on $V(\mathcal{F}, w_0, w)$ are given for the different function classes defined above. As it turns out, these upper bounds can then be minimized using quadratic programming (QP), yielding optimal (in this sense) estimators.

3 The DWO approach

3.1 Class $\mathcal{F} = \mathcal{F}_2(L, \delta, \Delta)$

Let us again consider the affine estimator (9) and the function class $\mathcal{F}_2(L, \delta, \Delta)$ for finite δ, Δ. For this estimator and class, the worst-case MSE (10) has the following upper bound:

$$\begin{aligned} &V(\mathcal{F}_2(L, \delta, \Delta), w_0, w) \le U_0(w_0, w) \\ &= \left(\left| w_0 + a \left(\sum_{k=1}^{N} w_k - 1 \right) + b \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \right| + \delta \left| \sum_{k=1}^{N} w_k - 1 \right| \right. \\ &\quad \left. + \Delta \left| \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \right| + \frac{L}{2} \sum_{k=1}^{N} |w_k| \widetilde{\varphi}_k^2 \right)^2 + \sigma^2 \sum_{k=1}^{N} w_k^2 \end{aligned} \tag{12}$$

This is true, since for any function $f \in \mathcal{F}_2(L, \delta, \Delta)$ the estimation error may be represented as follows

$$\begin{aligned} \hat{f}(\varphi_0) - f(\varphi_0) &= w_0 + a \left(\sum_{k=1}^{N} w_k - 1 \right) + b \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \\ &+ (f(\varphi_0) - a) \left(\sum_{k=1}^{N} w_k - 1 \right) + (f'(\varphi_0) - b) \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \\ &+ \sum_{k=1}^{N} w_k (f(\varphi_k) - f(\varphi_0) - f'(\varphi_0) \widetilde{\varphi}_k) + \sum_{k=1}^{N} w_k e_k \end{aligned} \tag{13}$$

Due to a well known lemma, the inequality

$$|f(\varphi_k) - f(\varphi_0) - f'(\varphi_0) \widetilde{\varphi}_k| \le \frac{L}{2} \widetilde{\varphi}_k^2 \tag{14}$$

follows from (2). Now, the MSE (11) satisfies

$$\begin{aligned} MSE(f, w_0, w) \le &\left(\left| w_0 + a \left(\sum_{k=1}^{N} w_k - 1 \right) + b \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \right| \right. \\ &+ |f(\varphi_0) - a| \cdot \left| \sum_{k=1}^{N} w_k - 1 \right| + |f'(\varphi_0) - b| \cdot \left| \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \right| \\ &\left. + \sum_{k=1}^{N} |w_k| \cdot |f(\varphi_k) - f(\varphi_0) - f'(\varphi_0) \widetilde{\varphi}_k| \right)^2 + \sigma^2 \sum_{k=1}^{N} w_k^2 \end{aligned} \tag{15}$$

from which the upper bound (12) follows directly.

Note that the upper bound $U_0(w_0, w)$ is easily minimized with respect to w_0 for any $w \in \mathbf{R}^N$. Indeed,

$$\arg\min_{w_0} U_0(w_0, w) = -a \left(\sum_{k=1}^{N} w_k - 1 \right) - b \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \tag{16}$$

Thus, we arrive at the following consequence: For the function class $\mathcal{F}_2(L, \delta, \Delta)$, the affine estimator (9) minimizing $U_0(w_0, w)$ may be sought among the estimators satisfying

$$\hat{f}(\varphi_0) = \sum_{k=1}^{N} w_k y_k - a \left(\sum_{k=1}^{N} w_k - 1 \right) - b \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \,, \; w \in \mathbf{R}^N \tag{17}$$

For this kind of estimators, $V(\mathcal{F}_2(L,\delta,\Delta),w_0,w)$ has the following upper bound, which follows directly from (12) and (16):

$$V(\mathcal{F}_2(L,\delta,\Delta),w_0,w) \le U_0(w)$$
$$= \left(\delta\left|\sum_{k=1}^N w_k - 1\right| + \Delta\left|\sum_{k=1}^N w_k\widetilde{\varphi}_k\right| + \frac{L}{2}\sum_{k=1}^N |w_k|\widetilde{\varphi}_k^2\right)^2 + \sigma^2\sum_{k=1}^N w_k^2 \quad (18)$$

Let us take a closer look at (13), and particularly the term

$$(f(\varphi_0)-a)\left(\sum_{k=1}^N w_k - 1\right)$$

It is easy to see, that if $\delta \to \infty$, this term is unbounded – and hence the MSE might be arbitrarily large – unless we have $\sum_{k=1}^N w_k = 1$. In fact, we can show the following theorem:

Theorem 3.1. *Assume that $\widetilde{\varphi}_k \neq 0$, $k = 1,\dots,N$. Given $a,b \in \mathbf{R}$ and $L,\Delta \in (0,+\infty)$, there exists a $\delta_0 \in (0,+\infty)$ such that for any $\delta > \delta_0$, the minimum of the upper bound $U_0(w)$ given by* (18) *with respect to $w \in \mathbf{R}^N$ is attained on the subspace*

$$\sum_{k=1}^N w_k = 1 \quad (19)$$

and does not depend on a or δ. In other words, given a sufficiently large δ, the affine estimator (9) *minimizing $U_0(w_0,w)$ can be found in the form*

$$\hat{f}(\varphi_0) = \sum_{k=1}^N w_k y_k - b\sum_{k=1}^N w_k\widetilde{\varphi}_k\,, \quad \sum_{k=1}^N w_k = 1 \quad (20)$$

with the weights $(w_1,\dots,w_N)$ minimizing the simpler upper bound

$$U_1(w) = \left(\Delta\left|\sum_{k=1}^N w_k\widetilde{\varphi}_k\right| + \frac{L}{2}\sum_{k=1}^N |w_k|\widetilde{\varphi}_k^2\right)^2 + \sigma^2\sum_{k=1}^N w_k^2 \quad (21)$$

subject to the constraint (19).

Proof. The proof is found in [7], and is based on assuming that (19) is not satisfied. Under this assumption, one can explicitly construct a δ_0, such that for any $\delta > \delta_0$, the minimal value of (18) is strictly larger than the minimum of (21) subject to the constraint (19). ☐

3.2 Class $\mathcal{F} = \mathcal{F}_2(L,\Delta)$

Let us now turn to class $\mathcal{F}_2(L,\Delta)$, i.e., the case $\delta \to \infty$. From the remark just before Theorem 3.1, it follows that the MSE cannot be bounded above unless $\sum_{k=1}^N w_k = 1$. On the other hand, if this requirement is satisfied, we get the following upper bound on the worst-case MSE, which can be shown analogously to (12):

$$V(\mathcal{F}_2(L,\Delta),w_0,w) \le U_1(w_0,w)$$
$$= \left(\left|w_0 + b\sum_{k=1}^N w_k\widetilde{\varphi}_k\right| + \Delta\left|\sum_{k=1}^N w_k\widetilde{\varphi}_k\right| + \frac{L}{2}\sum_{k=1}^N |w_k|\widetilde{\varphi}_k^2\right)^2 + \sigma^2\sum_{k=1}^N w_k^2 \quad (22)$$

By minimizing the upper bound with respect to w_0, it can be seen that the minimizing estimator will be in the form (20), and that w can be found by minimizing (21) under assumption (19).

Analogously to Theorem 3.1, we can study what happens when Δ is large.

Theorem 3.2. *Suppose that $\widetilde{\varphi}_k \neq 0$, $k = 1,\dots,N$, and that there are two indices k_1 and k_2 such that $\widetilde{\varphi}_{k_1} \neq \widetilde{\varphi}_{k_2}$. Given $b \in \mathbf{R}$ and $L \in (0,+\infty)$, there exists a $\Delta_0 \in (0,+\infty)$ such that for any $\Delta > \Delta_0$, the minimum of the upper bound $U_1(w)$ given by* (21)*, subject to the constraint* (19)*, is attained on the subspace*

$$\sum_{k=1}^N w_k = 1\,, \quad \sum_{k=1}^N w_k\widetilde{\varphi}_k = 0 \quad (23)$$

and does not depend on b or Δ. In other words, given a sufficiently large Δ, the affine estimator (9) *minimizing $U_1(w_0,w)$ can be found in the form* (8) *by minimizing the upper bound*

$$U_2(w) = \left(\frac{L}{2}\sum_{k=1}^N |w_k|\widetilde{\varphi}_k^2\right)^2 + \sigma^2\sum_{k=1}^N w_k^2 \quad (24)$$

subject to constraints (23).

Proof. Similar to the proof of Theorem 3.1. See [7]. ☐

3.3 Class $\mathcal{F} = \mathcal{F}_2(L)$

For class $\mathcal{F}_2(L)$, following a similar line of argument as for $\mathcal{F}_2(L,\Delta)$, we can see that a finite MSE can be guaranteed only if the weights w satisfy (23). Under this requirement, on the other hand, we get the following upper bound on the worst-case MSE:

$$V(\mathcal{F}_2(L),w_0,w) \le U_2(w_0,w)$$
$$= \left(|w_0| + \frac{L}{2}\sum_{k=1}^N |w_k|\widetilde{\varphi}_k^2\right)^2 + \sigma^2\sum_{k=1}^N w_k^2 \quad (25)$$

Hence, by minimizing the upper bound with respect to w_0, we will obtain a minimizing estimator in the form (8), where w can be found by minimizing (24) subject to the constraints (23).

4 QP formulation

In Section 3, it was pointed out how the weights w_0 and w_k of the affine estimator (9) could be chosen by minimizing different expressions, in order to get estimators with a guaranteed upper bound on the worst-case MSE. In this section we will show that these minimization problems can be formulated as *convex quadratic programs (QP)* (see [1]).

To begin with, let us consider the function class $\mathcal{F}_2(L, \delta, \Delta)$ and the problem of finding the affine estimator minimizing (18).

Theorem 4.1. *Given the positive numbers δ, Δ, L, consider the following minimization problem:*

$$\begin{aligned} \min_{w,s} \quad & \left(\delta s_a + \Delta s_b + \frac{L}{2}\sum_{k=1}^{N} \widetilde{\varphi}_k^2 s_k\right)^2 + \sigma^2 \sum_{k=1}^{N} s_k^2 \\ \text{subj. to} \quad & s_a \geq \pm\left(\sum_{k=1}^{N} w_k - 1\right) \\ & s_b \geq \pm \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \\ & s_k \geq \pm w_k, \qquad k = 1, \dots, N \end{aligned} \tag{26}$$

where $w = (w_1, \dots, w_N)$ and $s = (s_a, s_b, s_1, \dots, s_N)$. Then w^* is a minimizer of (18) *if and only if there is a vector s^* such that (w^*, s^*) is a minimizer of* (26)*.*

Proof. Given a feasible solution w to (18), we can get a feasible solution to (26) with the same value of the objective function by using the same w and

$$\begin{aligned} s_a &= \left|\sum_{k=1}^{N} w_k - 1\right| \\ s_b &= \left|\sum_{k=1}^{N} w_k \widetilde{\varphi}_k\right| \\ s_k &= |w_k|, \qquad k = 1, \dots, N \end{aligned} \tag{27}$$

Hence (26) is a relaxation of (18), and it suffices to show that for a minimizer (w^*, s^*) of (26), (27) will hold. Suppose, e.g., that $s_1^* > |w_1^*|$. 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^* is non-negative in the first sum of the objective function, and positive in the second sum of the objective function, so decreasing s_1^* will decrease at least one of these sums, and hence the objective function. Hence, $s_1^* = |w_1^*|$. By similar arguments, one can show that the other equalities of (27) will also hold at the optimum, and the theorem is proven. ☐

Remark 4.1. Note that (26) is a convex QP and can therefore be solved efficiently.

Starting from Theorem 4.1, we can now formulate QP:s for all the other cases mentioned in Section 3. Since the constraints (19) and (23) are all linear in w, they can just be added to the QP.

5 Estimating multivariate functions

In this paper, we have so far 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. In this section, we will describe some of the aspects of this kind of extension. In the following, $\|\cdot\|$ will denote the Euclidean norm, and b^T denotes the transpose of b.

We now consider the problem of estimating the value $f(\varphi_0)$ of an unknown multivariate, continuously differentiable function $f : \mathbf{R}^n \to \mathbf{R}$ at a given point φ_0, given a set of input-output pairs $\{(\varphi_k, y_k)\}_{k=1}^N$, coming from the relation

$$y_k = f(\varphi_k) + e_k \tag{28}$$

Instead of the assumptions (2)-(4), we make the following assumptions:

$$\|\nabla f(\varphi_1) - \nabla f(\varphi_2)\| \leq L\|\varphi_1 - \varphi_2\| \quad \forall\, \varphi_1, \varphi_2 \in \mathbf{R}^n \tag{29}$$

$$|f(\varphi_0) - a| \leq \delta\,, \tag{30}$$

$$\|\nabla f(\varphi_0) - b\| \leq \Delta\,. \tag{31}$$

With some abuse of notation, we let $\mathcal{F}_2(L, \delta, \Delta)$, $\mathcal{F}_2(L, \Delta)$, and $\mathcal{F}_2(L)$ denote also their multivariate counterparts. For $\mathcal{F}_2(L, \delta, \Delta)$ and an affine estimator (9), the worst-case MSE (10) has the following upper bound, which is similar to the univariate case:

$$\begin{aligned} & V(\mathcal{F}_2(L, \delta, \Delta), w_0, w) \leq U_0(w_0, w) \\ &= \Bigg(\left|w_0 + a\left(\sum_{k=1}^{N} w_k - 1\right) + b^T \sum_{k=1}^{N} w_k \widetilde{\varphi}_k\right| \\ &\quad + \delta\left|\sum_{k=1}^{N} w_k - 1\right| + \Delta\left\|\sum_{k=1}^{N} w_k \widetilde{\varphi}_k\right\| + \frac{L}{2}\sum_{k=1}^{N} |w_k| \|\widetilde{\varphi}_k\|^2\Bigg)^2 \\ &\quad + \sigma^2 \sum_{k=1}^{N} w_k^2 \end{aligned} \tag{32}$$

As in the univariate case, we can immediately eliminate w_0 by minimizing (32) for an arbitrary w, giving that the affine estimator minimizing $V(\mathcal{F}_2(L, \delta, \Delta), w_0, w)$ may be sought among the estimators satisfying

$$\hat{f}(\varphi_0) = \sum_{k=1}^{N} w_k y_k - a\left(\sum_{k=1}^{N} w_k - 1\right) - b^T \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \tag{33}$$

For this kind of estimators, the worst-case MSE (10) over the class $\mathcal{F}_2(L, \delta, \Delta)$ has the following upper bound:

$$\begin{aligned} & V(\mathcal{F}_2(L, \delta, \Delta), w_0, w) \leq U_0(w) \\ &= \left(\delta\left|\sum_{k=1}^{N} w_k - 1\right| + \Delta\left\|\sum_{k=1}^{N} w_k \widetilde{\varphi}_k\right\| + \frac{L}{2}\sum_{k=1}^{N} |w_k| \|\widetilde{\varphi}_k\|^2\right)^2 \\ &\quad + \sigma^2 \sum_{k=1}^{N} w_k^2 \end{aligned} \tag{34}$$

So far, the differences to the univariate case have been small and obvious. However, when trying to transform the problem of minimizing (34) into a standard convex optimization problem, it turns out that it is impossible to formulate it as a QP problem. What prohibits this is the term

$$\Delta \left\| \sum_{k=1}^{N} w_k \widetilde{\varphi}_k \right\| \tag{35}$$

which is the norm of a linear combination of vectors. Instead, we can formulate the problem as a *second-order cone program (SOCP)*, which is another standard class of convex optimization problems (see [1]). To do this, we introduce some slack variables $s = (s_1 \ \dots \ s_N)^T$ and $t = (t_a \ t_b \ t_c)^T$, and get

$$\begin{aligned} &\min_{w,s,t} t_c \\ &\text{subj. to } \left(\delta t_a + \Delta t_b + \frac{L}{2}\sum_{k=1}^{N} \|\widetilde{\varphi}_k\|^2 s_k\right)^2 + \sigma^2 \sum_{k=1}^{N} s_k^2 \le t_c \\ &\qquad \left|\sum_{k=1}^{N} w_k - 1\right| \le t_a \\ &\qquad \left\|\sum_{k=1}^{N} w_k \widetilde{\varphi}_k\right\| \le t_b \\ &\qquad |w_k| \le s_k, \qquad k = 1, \dots, N \end{aligned} \tag{36}$$

This problem is in standard SOCP form, except for the first, quadratic constraint. However, straightforward calculations show that this constraint is equivalent to

$$\left\| \begin{pmatrix} 2\left(\delta t_a + \Delta t_b + \frac{L}{2}\sum_{k=1}^{N} \|\widetilde{\varphi}_k\|^2 s_k\right) \\ 2\sigma s \\ 1 - t_c \end{pmatrix} \right\| \le 1 + t_c \tag{37}$$

thus completing the problem reformulation.

For the other function classes, $\mathcal{F}_2(L, \Delta)$ and $\mathcal{F}_2(L)$, the extension to the multivariate case is done completely similarly. The minimization problem for $\mathcal{F}_2(L, \Delta)$ will also yield a SOCP, while the minimization problem for $\mathcal{F}_2(L)$ will still be possible to express as a QP, since the term (35) vanishes.

6 Example

The following simple example illustrates the fact that the information about bounds on the function value and derivatives can be useful, but only if the bounds are tight enough.

Let us consider the simple nonlinear system (with $f : \mathbf{R}^2 \to \mathbf{R}$)

$$\begin{aligned} y_k &= f(\varphi_k) + e_k \\ f(\varphi_k) &= 5(\varphi_{k1}^2 + \varphi_{k2}^2) + 5\varphi_{k1} + 10\varphi_{k2} + 15 \end{aligned} \tag{38}$$

where $e_k \in N(0, 1)$. Suppose that we would like to estimate y for $\varphi_0 = \begin{pmatrix} 0 & 0 \end{pmatrix}^T$, given only $N = 20$ data samples taken from the distribution $\varphi_k \in N(0, I)$. We assume that we know that (29) is satisfied for $L = 10$ (this is, of course, the best possible L), and that we also know some approximate values a and b of the function $f(\varphi_0)$ and $\nabla f(\varphi_0)$, respectively, and the bounds δ and Δ according to (30) and (31). For simplicity, let a and b be the true values, $a = f(\varphi_0)$ and $b = \nabla f(\varphi_0)$.

Now we can use the estimator in (33), for which the appropriate weights are obtained by solving the SOCP given by (36) and (37). Solving the SOCP can be done using YALMIP [4] and SEDUMI [13]. Naturally, different δ and Δ values should give different estimates. Figure 1(a) shows $\hat{f}(\varphi_0)$ for some different values of δ and Δ. In Figure 1(b) the part of the estimate coming from the a priori knowledge of the function value (i.e., the second term of (33)) is plotted. As we can see, for small values of δ, the estimate is based entirely on this information, while for large values, the a priori knowledge is not used at all, in agreement with Theorem 3.1. In Figure 1(c), the part of the $\hat{f}(\varphi_0)$ coming from the knowledge of $\nabla f(\varphi_0)$ (the third term of (33)) is used. For this example, we can see that this information is not used very much, but that it gives a certain contribution for small values of Δ (as long as δ is large enough, so that we do not only use the information about $f(\varphi_0)$).

Finally, Figure 2 shows the optimal value of the criterion function, which decreases both with decreasing δ and Δ, just as should be expected.

7 Conclusions

In this paper, a direct weight optimization approach for local modelling has been presented. It was shown that the approach is easily extended to the case when a priori bounds on the function and its derivative are known. Theorems 3.1 and 3.2 however showed that when the bounds are very wide, the extra information may not be enough to improve the solution.

The application of these methods to dynamical systems with the regressors φ_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 [12] for such ideas.

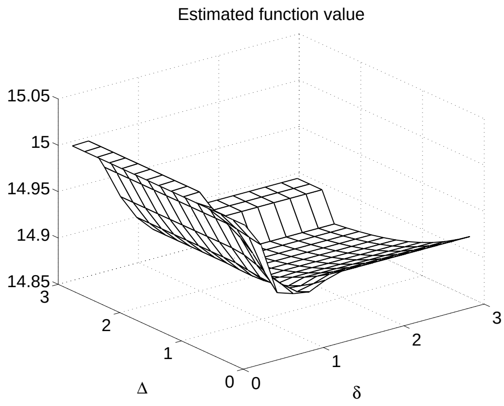
References

[1] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Course Reader for EE364, Introduction to Convex Optimization with Engineering Applications, Stanford University, May 3, 1999.

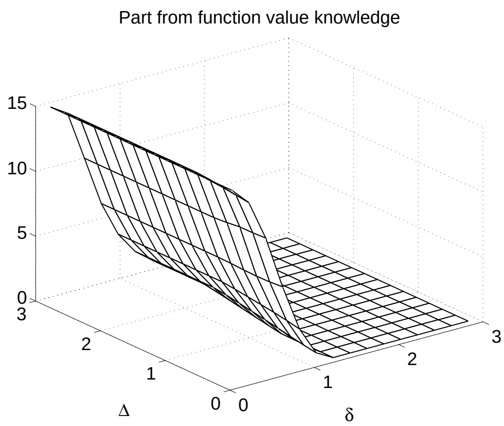
[2] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification And Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

[3] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.

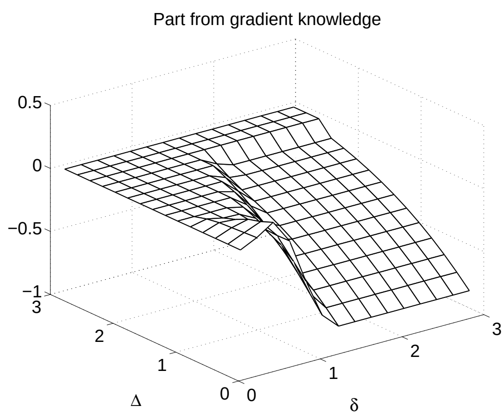
[4] Johan Löfberg. *YALMIP 2.2 – Yet Another LMI Parser*. Linköping University, SE-581 83 Linköping, Sweden, 2002. `http://www.control.isy.liu.se/~johanl/yalmip.html`.



(a) The estimate $\hat{f}(\varphi_0)$.



(b) The part of $\hat{f}(\varphi_0)$ coming from prior knowledge of $f(\varphi_0)$.



(c) The part of $\hat{f}(\varphi_0)$ coming from prior knowledge of $\nabla f(\varphi_0)$.

Figure 1: Estimates of $f(\varphi_0)$ in the example for different values of δ and Δ.

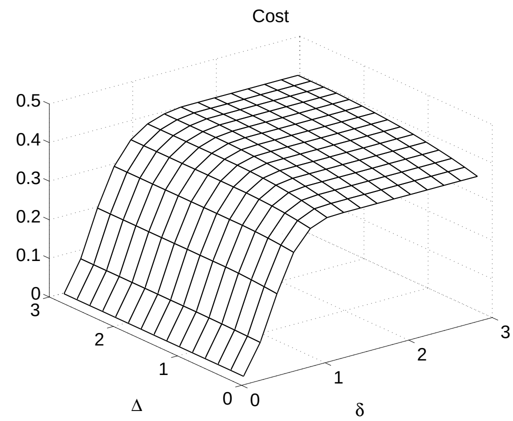


Figure 2: The optimal value of the criterion function in the example.

[5] Roderick Murray-Smith and Tor Arne Johansen, editors. *Multiple Model Approaches to Modelling and Control*. Taylor & Francis, 1997.

[6] E. A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 10:186–190, 1964.

[7] Jacob Roll. *Local and Piecewise Affine Approaches to System Identification*. PhD thesis, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, April 2003.

[8] Jacob Roll, Alexander Nazin, and Lennart Ljung. A non-asymptotic approach to local modelling. In *The 41st IEEE Conference on Decision and Control*, pages 638–643, December 2002.

[9] Jacob Roll, Alexander Nazin, and Lennart Ljung. Local modelling of nonlinear dynamic systems using direct weight optimization. Accepted for the 13th IFAC Symposium on System Identification, Rotterdam, August 2003.

[10] J. Sacks and D. Ylvisaker. Linear estimation for approximately linear models. *The Annals of Statistics*, 6(5):1122–1137, 1978.

[11] 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*, 31(12):1691–1724, 1995.

[12] Anders Stenman. *Model on Demand: Algorithms, Analysis and Applications*. PhD thesis, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden, 1999.

[13] Jos F. Sturm. *SeDuMi 1.05*, 2002. `http://fewcal.kub.nl/sturm/software/sedumi.html`.

[14] Geoffrey S. Watson. Smooth regression analysis. *Sankhyā, Series A*, 26:359–372, 1964.