Non-Linear Dynamics Isolated by Delaunay Triangulation Criteria

David Lindgren and 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: david@isy.liu.se

March 2, 2004

Report no.: LiTH-ISY-R-2598
Submitted to The 2004 IEEE Conference on Decision and Control

Technical reports from the Control & Communication group in Linköping are available at http://www.control.isy.liu.se/publications.
Non-Linear Dynamics Isolated by Delaunay Triangulation Criteria

David Lindgren and Lennart Ljung

March 2, 2004

Abstract

We show that Delaunay triangulation of data points sampled from a system with an additive non-linearity gives a criterion by which a linear projection can be found that isolates the non-linear dependence, leaving out the linear one. This isolation means the non-linear modeling can be confined to a regressor space of lower dimensionality, which in turn means over-parameterization can be avoided. Grassmann manifold optimization techniques specialized for finding linear projections are used to minimize the triangle criteria. Monte Carlo simulations indicate that a particular criterion built on triangle asymmetries has a minimum that coincides with the sampled system non-linear part. The criterion is however complex to compute and non-convex, which makes it difficult to optimize globally.

1 Introduction

We consider the modeling of a time-discrete SISO-system,

\[ y_t = f(\phi_t) + v_t \]

\[ \phi_t = \begin{bmatrix} y_{t-1} & y_{t-2} & \cdots & y_{t-n_a} & u_t & u_{t-1} & \cdots & u_{t-n_b} \end{bmatrix}^T \]  

(1)

where we know, or have reason to believe, that the non-linearities in \( f(\cdot) \) are restricted to a Euclidean linear subspace of the regressor,

\[ f(\phi_t) = g(A^T \phi_t) + B^T \phi_t. \]  

(2)

Here, \( u_t \) is the system input, \( y_t \) the output and \( v_t \) uncertainties. \( A \) is an \( n \)-by-\( k \) matrix with full column rank and \( k < n = n_a + n_b + 1 \). The column space of \( A \) identifies the linear subspace to which the non-linearities are confined. \( B \) in \( \mathbb{R}^n \) accounts for the linear dependencies. A standing assumption is that the linear subspaces of \( A \) and \( B \) are linearly independent, \( A^T B = 0 \).

The points \( y_t - v_t, t = 1, 2, \ldots, N \), lie on a hyper-surface over \( n \) dimensions. Modeling the system is the same as modeling this surface. Of course, it is...
difficult to visualize and conceptualize this surface when \( n > 2 \). However, the points on the “subsurface”

\[
y_t - v_t = \tilde{g}(A^T \phi_t)
\]

over \( k \) dimensions allow us to isolate the non-linear dependence. This is thanks to the linear independence of \( A \) and \( B \). If \( k < n \), the non-linear modeling problem is greatly simplified by this projection. Note that we can only achieve this since the residual \( B^T \phi_t \) is linear and lies in the linear null-space of \( A \). In words, the linear contribution is projected away.

To find \( A \) seems to be a natural first step in the modeling of a system of this type. How to do this is the main topic of this work. We assume a data set with samples \( \{y_t, u_t\}_1^N \) is available. Below is described a specific criterion based on the triangles of a sample Delaunay triangulation, see Fig. 1 and below. The criterion forms a projection index or an objective function that we will show has excellent selectivity in respect of separating the space spanned by the columns of \( A \) and \( B \). However, the objective function is non-convex and the global optimum may be very hard to attain.

### 1.1 Example: The Non-Linear Regressors of a Tank Process

A drained tank with water level \( y(t) \) and inlet flow \( u(t) \) can be described by

\[
\dot{y}(t) = u(t) - \sqrt{y(t)} + n(t).
\]
Here, \( n(t) \) accounts for uncertainties. If the tank is sampled with interval \( T \), we would get a sequence \( \{y_t = y(tT)\}_1^N \) that is rather well described by

\[
\Delta y_t = \left( y_{t+1} - y_t \right)/T = u_t - \sqrt{y_t} + v_t. \tag{5}
\]

If we now give the sampled sequence bundled with \( \{u_t\}_1^N \) to a user not telling the user from where it originates, and with the task of finding a black box model \( f \) that makes the expected magnitude of the residual

\[
e_t = f(\Delta y_t, y_t, u_t)^T = f(\phi_t) \tag{6}
\]

small, the user’s first attempt to take \( f \) linear would of course be unsatisfactory. However, when the user plots the sample points in 3 dimensions \( \{\Delta y_t, y_t, u_t\} \) and rotate them interactively to study different projections, the user finds that for a certain projection,

\[
\phi_{\text{null}} = \begin{bmatrix} a_1^T \phi_t \\ a_2^T \phi_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \phi_t \tag{7}
\]

a dependence between two directions \( a_1 \) and \( a_2 \) in the regressor space is evident, see Fig. 2. In fact, what the user sees in the plot is the non-linear residual of a linear model, and from this view it is rather clear that

\[
f(\phi_t) = a_2^T \phi_t - \sqrt{a_1^T \phi_t} \tag{8}
\]

indeed is an adequate model.

Figure 2: Isolated, non-linear regressor of a sampled tank process. The linear contribution lies in the null-space of this projection. Given this plot, the step to an adequate non-linear model is short.
1.2 Related Work

From the modeler’s perspective, visual feedback from software packages for exploratory data analysis like XGobi is very valuable, see [2, 3]. Here, different projections of the regressor may be pursued interactively and structures and non-linearities can be detected and subsequently modeled, exploratory projection pursuit, see [4, 5, 6]. There is also measure of frustration involved in using these tools due to our difficulties to conceptualize high-dimensional regressor spaces, even more so, to navigate in them. Therefore a number of projection indices have been proposed to automate the pursuit, and leave to an algorithm the search through high-dimensional spaces for interesting projections, see [7, 8, 9, 10]. These indices or criteria are based on entropy, non-normality, cluster structure etc. See also the textbook [11].

This work is closely related to projection pursuit. What we (up to now) have lacked, though, is a projection index that finds projections where the structure of non-linearities is revealed. Such an index would be valuable both for visualization and parameter estimation.

2 Delaunay Triangulation

Delaunay triangulation is a way to connect points with straight lines. The connections form triangles such that no data point falls in any triangle’s circumscribed circle, see [12]. [13] proposes a program that performs Delaunay triangulation of a set of \( N \) points with time complexity \( O(N \log N) \). An implementation for Delaunay triangulation is included as a standard polynomial function in MATLAB, \texttt{delaunayn}.

We will use a measure of non-linear dependence suggested by [14], who models a set of \( N \) points \( x_t \) in \( \mathbb{R}^k \) as

\[
 f(x_t) = 0,
\]

where \( f \) is not necessarily invertible (regression is a special case). [14] uses a Delaunay triangulation of the data points, and shows in a simulation that two particular functionals of the triangle set are related to the dependence between the entries of \( x_t \) in \( \mathbb{R}^2 \). The two functionals build on

1. Triangle area. If the dependence between entries in \( x_t \) is strong, the majority of the triangles have very small area compared to the convex hull. The criterion is

\[
 F_a = \frac{T_r - T_h + T_1 + T_2 + \cdots + T_p}{T_r},
\]

where \( T_r \) is the area of a rectangular region defined by the extreme points, \( T_h \) the area of the convex hull, and \( T_1 + T_2 + \cdots + T_p \) the area of the \( p \) triangles with smallest area. \( p \) is chosen about half the number of triangles.

2. Triangle asymmetry. If the dependence between entries in \( x_t \) is large, the triangles are thin (or asymmetric). The criterion is

\[
 F_s = -\frac{1}{m^2} \sum_{i=1}^{m} \frac{f^{(i)}_{1}}{l_{1}^{(i)}} + \frac{f^{(i)}_{2}}{l_{2}^{(i)}} + \frac{f^{(i)}_{3}}{l_{3}^{(i)}},
\]
where \( m \) is the number of triangles, \( l_j^{(i)} \) the length of side \( j \) in triangle \( i \). The sides are ordered according to \( l_1^{(i)} \leq l_2^{(i)} \leq l_3^{(i)} \).

It is here motivated to offer an intuitive explanation to why we expect these criteria to do the job. The plane, perpendicular projection of a set of parallel lines is obviously a set of points. This is not the case for non-linear curves. The contribution from the linear part are samples lying on fictitious lines, and again, a projection perpendicular to these lines, collects the sample points on a minimal area. In fact, the data points are, up to the noise term, collected exactly on the non-linear curve. This motivates immediately the area criterion above. Furthermore, if we do Delaunay triangulation on points that are all close to a curve, the triangles will be very thin or asymmetric.

3 Finding \( A \) by Optimization

By the column space of the \( n \)-by-\( k \) matrix \( S \) we represent the linear subspace where we seek to confine the non-linearity. The projected data point is of course a function of the projection \( S \), and the triangle functionals will be based on the data set

\[
x_t(S) = \begin{bmatrix} y_t \\ S^T \phi_t \end{bmatrix}, \quad t = 1, 2, \ldots, N.
\]

(12)

This implies in turn that also the functionals (10, 11) above form (objective) functions of \( S \): \( F_a(S) \) and \( F_s(S) \). In a simulation below, we are interested in the subspaces obtained minimizing these objective functions,

\[
S^*_a = \arg \min_S F_a(S), \quad \text{s.t.} \quad S^T S = I
\]

(13)

and

\[
S^*_s = \arg \min_S F_s(S), \quad \text{s.t.} \quad S^T S = I.
\]

(14)

The constraint \( S^T S = I \) ensures the full rank (and good condition number) of \( S \).

Below we will address two major issues by Monte Carlo simulations:

1. Is the subspace spanned by \( S^*_a \) or \( S^*_s \) the same or almost the same as the one spanned by \( A \)?

2. Are the optimization problems tractable?

We will assume that the subspace dimensionality \( k \) is known beforehand.

3.1 Optimization Techniques

First we note that the lengths \( l_j^{(i)} \) in \( F_a(S) \) and the areas \( T_j \) in \( F_s(S) \) are invariant to orthonormal transformations \( Q \), why \( F_a(S) = F_a(Q^T S) \) and \( F_s(S) = F_s(Q^T S) \) for every \( Q \) with \( Q^T Q = I \). This property is denoted homogeneity. The homogeneity together with the orthonormal constraint \( S^T S = I \) characterizes the problem as optimization over the Grassmann manifold. The state of the art algorithm for this is presented in [15], the non-linear conjugate gradients over the Grassmann manifold. This algorithm requires the gradient, \( \nabla F(S) \).
which can be analytically derived rather easily for at least (11). For our initial studies we will however settle with a simpler optimization technique. We will parameterize $S$ with Givens rotations.

The Givens rotation, defined as a plane rotation in $n$ dimensions, is used in QR factorization, see [16]. As we know it, it has not been used very extensively for what we here denote subspace optimization. One application, though, is the Optimal Discriminative Projection, which targets the classification accuracy in high-dimensional (multicollinear) data sets, see [17]. A related method that should be mentioned was reported by [18], who parameterized a linear combination of two of Fisher’s linear discriminants by an angle.

**Definition 1 (Givens Projection)** For a real parameter vector $p$ composed as

$$p = \begin{bmatrix} v \\ \bar{p} \end{bmatrix},$$

where $\dim v = n - 1$ and $\dim p = k(2n - k - 1)/2$, define the Givens projection from $\mathbb{R}^n$ to $\mathbb{R}^k$ as

$$S_{nk}(p) = G_n^T(p) \begin{bmatrix} e_1 & e_2 & \cdots & e_k \end{bmatrix},$$

$$G_n(p) = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & G_{n-1}(\bar{p}) \end{bmatrix} & \text{dim } p > n - 1 \\ r_n(p) & \text{dim } p = n - 1 \end{cases},$$

$$r_n(v) = \begin{cases} \begin{bmatrix} r_{n-1}(\bar{v}) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I_{n-2} & 0 \\ 0 & r_2(v_1) \end{bmatrix} & n > 2 \\ \begin{bmatrix} \cos v_1 & -\sin v_1 \\ \sin v_1 & \cos v_1 \end{bmatrix} & n = 2 \end{cases}.$$

Here, $e_i$ is a vector with zeros except for entry $i$, which is 1. It can be shown that $S_{nk}(p)S_{nk}(p) = I$ for every parameter vector $p$. Furthermore, for every $k$-dimensional subspace of $\mathbb{R}^n$, it exist a parameter vector $p$ such that the columns of $S_{nk}(p)$ span exactly this subspace, see [17].

Using the Givens projection casts the optimization problem constrained by orthonormality into an unconstrained one with $k(2n - k - 1)/2$ parameters;

$$p^* = \arg\min_p F(S_{nk}(p))$$

$$S^* = S_{nk}(p^*).$$

This problem can be solved by standard programs available for unconstrained minimization.

### 4 Simulation Study

By simulations we will investigate how $S_{a}^*$ and $S_{s}^*$ relate to $A$, and how complex the optimization problems (13) and (14) are. For this purpose we will use a slightly idealized setting. We will perturb a non-linear curve (surface) in $k = 2$ dimensions with noise, and augment the 2 dimensions with additional dimensions representing the linear part of the system. This augmented part we fill with white noise. The justification for using noise instead of an explicit
linear model for the augmented part, is that unmodeled dependencies can very well appear noise-like. For the optimization, we use the Givens projection described above together with the MATLAB function \texttt{fminunc} for unconstrained minimization.

### 4.1 Non-Linearities

We take \( N \) samples from two different curves (representing two modeling cases) parameterized in \( \xi \):

\[
x_t = \begin{bmatrix} \sin \pi \xi_t \\ \sin 1.7 \xi_t \end{bmatrix}
\]

(20)

and

\[
x_t = \begin{bmatrix} 1 - \xi^2 \\ \xi_t \frac{1 - \xi_t^2}{1 + \xi_t^2} \end{bmatrix}
\]

(21)

\( \xi \) is taken uniformly random, \(-1.5 < \xi < 1.5\). Note that \( \xi \) is an uninteresting variable, we just use it to define the curves. More important, \( \xi \) is not observed and will not help the modeler. Note also, that these curves are not in any way typical for system identification, but indeed non-trivial to model. In addition, (21), which is due to [14], is non-invertible at the origin. Samples are depicted in Fig. 3 and Fig. 6.

Now, \( N \) samples from a normal distribution with zero mean and covariance matrix \( \lambda_1 I_2 \), \( N(0, \lambda_1 I_2) \), are added to \( x_t \), and the points are connected by Delaunay triangulation. For \( \lambda_1 = 0.01 \), the results are depicted in Fig. 4 and Fig. 1. By increasing the noise variance \( \lambda_1 \), it becomes more difficult to “see” and make inferences about the relationship between the the two components of \( x_t \), see Fig. 5 and Fig. 7. For low noise variance \( \lambda_1 \), we note that:
Figure 4: Samples from (20) perturbed by noise samples from the the distribution $N(0, 0.01I)$. The sample points are connected by a Delaunay triangulation.

Figure 5: Samples from (20) perturbed by noise from the the distribution $N(0, 0.03I)$. The sample points are connected by a Delaunay triangulation. Compared to Fig. 4, the noise has larger variance and the triangles are more symmetric.
Figure 6: Samples from (21) with no noise added.

Figure 7: Delaunay triangulation of samples from (21) perturbed by noise from the normal distribution $N(0, 0.03I)$. Compared to Figure 1, the noise has larger variance and the triangles are more symmetric.
Figure 8: Iso-curves of $F_s(S_{32}([p_1 p_2]^T))$. 50 random data points on the curve (21) are mixed with noise, $\lambda_1 = 0.02$, $\lambda_2 = 0.3$. The cross in the middle accounts for the ideal point $S = A$. The minimizer of $F_s(S)$ is apparently very close to $A$.

- Few large triangles accounts for a very large portion of the area of the convex hull.
- The major part of the triangles are very thin or asymmetric.

These observations agrees well with the definitions of $F_a$ and $F_s$, respectively.

4.2 High-dimensional Augmentation

As mentioned, the 2-dimensional vector $x_t$ is augmented by a random noise vector $z_t$ sampled from $N(0, \lambda_2 I_{n-2})$ thus forming

$$\bar{x}_t = \begin{bmatrix} x_t \\ z_t \end{bmatrix}$$

with $n = 7$ dimensions and consequently,

$$x_t = A^T \bar{x}_t, \quad \text{for } A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}.$$ (23)
Figure 9: Iso-curves of $F_\alpha(S_{32}([p_1 \ p_2]^T))$. 50 random data points on the curve (21) are mixed with noise, $\lambda_1 = 0.02, \lambda_2 = 0.3$. The cross in the middle accounts for the ideal point $S = A$. The minimizer of $F_\alpha(S)$ is not very close to $A$.

4.3 Objective Function Contour Plot for $n = 3, k = 2$

By the study of contour-plots or iso-curves of $F(S_{32}(p))$ (angle parameterized projections from 3 to 2 dimensions) we may gain apprehension of the objective function complexity. Here, we can also immediately see if the global minimizer is close to $A$.

The parameter space in this case is two dimensional, $p = [p_1 \ p_2]$, and $0 \leq p_i < 180^\circ$. 50 random data points on the curve (21) are mixed with noise, $\lambda_1 = 0.02$ and $\lambda_2 = 0.3$. The iso-curves of the objective function $F_\alpha$ are depicted in Fig. 8, and of $F_\alpha$ in Fig. 9. We learn that:

- $F_\alpha(S)$ and $F_s(S)$ are non-convex and have many local minima.
- The global minimum $S^*_\alpha$ is very close to the ideal point $S = A$ (for this special instance). This is not the case for $S^*_s$.

In Fig. 10 we depict the iso-curves of $F_s(S_{32}(p))$ for the sampled tank (simulated by SIMULINK) from the introductory example. This plot confirms what we have observed; $F_s$ induce a very attractive optimum, but the many local optima prevent us from finding it by local minimization.

4.4 Minimization with Start in $A$, Monte Carlo

We now investigate if $A$ is close to a (local) minimizer $\hat{S}$ of $F_\alpha(S)$ and $F_s(S)$. We conduct this experiment in $n = 7$ dimensions. The minimization starting point $S_0$ is given by $S_0 = A$. 

11
Figure 10: Iso-curves of $F_s(S_{32}([p_1\ p_2]^T))$. The data is sampled from a simulated water tank. Here, $A$ has the coordinate $p_1 = 90^\circ$, $p_2 = 135^\circ$, which is indistinguishable from the global minimizer.

Table 1: Results of 30 minimizations with different realizations and starting point $S_0 = A$. In each realization, 50 random data points on the curve (21) are mixed with noise, $\lambda_1 = 0.02$, $\lambda_2 = 0.3$. Projection from 7 to 2 dimensions.

|       | $F(S) - F(A)$, % | $d(S|A)$, % |
|-------|-----------------|-------------|
| $F_u(S)$ | $-5.80 \pm 5.12$ | $4.3 \pm 6.4$ |
| $F_s(S)$ | $-0.40 \pm 0.48$ | $0.02 \pm 0.035$ |

We now define the distance from a projection $S$ to the ideal point $A$ as

$$d(S|A) = \frac{1}{2} \sum_{j=1}^{7} \sum_{i=1}^{2} s^2_{ij}.\quad (24)$$

If $d(S|A) = 0$, then $S$ and $A$ spans the same subspace. On the other hand, if $d(S|A) = 1$, $S$ and $A$ are orthogonal.

The results for the curve (21) is given in Table 1 and for (20) in Table 2. The results are given as mean value $\pm$ standard deviation. We learn that in both cases, $A$ is very close to a (local) minimizer of $F_s(S)$. This is, however, not the case for $F_u(S)$.

4.5 Minimization with Random Start, Monte Carlo

The start $S_0$ is now a perturbation of $A$ in a random direction, but with a fixed $d(S_0|A)$ given as a simulation parameter.

The results for data from (21) is given in Table 3. If $d(S_0|A) = 5\%$ the minimization will almost always converge to $A$. If $d(S_0|A) \geq 10\%$, this is not the case. We learn that we cannot hope to reach $A$ unless we start the minimization very close to it.
Table 2: Results of 30 minimizations with different realizations and starting point $S_0 = A$. In each realization, 50 random data points on the curve (20) are mixed with noise, $\lambda_1 = 0.04$, $\lambda_2 = 0.4$. Projection from 7 to 2 dimensions.

|       | $F(S) - F(A)$, % | $d(S|A)$, % |
|-------|------------------|-------------|
| $F_0(S)$ | $-32.19 \pm 5.04$ | $6.51 \pm 10$ |
| $F_2(S)$ | $-0.61 \pm 0.90$ | $0.065 \pm 0.14$ |

Table 3: Results of 30 minimizations of $F_s(S)$ with different realizations and random starting point $S_0$ lying at different distances from $A$. In each realization, 50 random data points on the curve (21) are mixed with Gaussian noise $\lambda_1 = 0.02$, $\lambda_2 = 0.3$. Projection from 7 to 2 dimensions.

|       | $d(S_0|A)$, % | $F(S) - F(A)$, % | $F(S) - F(S_0)$ | $d(S|A)$, % |
|-------|---------------|------------------|-----------------|-------------|
| 5     | $0.0 \pm 0.03$ | $-0.08 \pm 0.03$ | $0.01 \pm 0.02$ |             |
| 10    | $1.78 \pm 2.57$ | $-6.16 \pm 2.78$ | $6.67 \pm 5.76$ |             |
| 20    | $4.97 \pm 2.56$ | $-5.5 \pm 2.29$ | $26 \pm 19$    |             |
| 80    | $5.22 \pm 2.87$ | $-6.8 \pm 3.0$  | $56 \pm 34$    |             |

4.6 Notes on Complexity

Starting randomly, it takes few seconds (3-10) for the minimization program to reach a local minimizer. The convergence is fairly good. If $k = 2$, it is no problem to solve problems with a few hundred sample points.

5 Conclusions and Future Works

5.1 Conclusions

The Delaunay triangulation offers a criterion that has potential to be very useful in system identification. The criterion is based on triangle asymmetries and has been found to have excellent selectivity between subspaces with non-linear dependence and subspaces with linear dependence. Furthermore, with special optimization techniques, for instance parameterization with Givens rotations, we can rather quickly find a linear subspace that is a local minimizer of the criterion. However, the criterion is non-convex with many local minima. For the criterion to be useful, we must either advance the optimization technique or find good start solutions that lie close to the global optimum.

5.2 Future Works

In the future we would like to investigate:

- Are there any optimization techniques that better can deal with the non-convexity? Simulated annealing?
- Are there any simple start solution close to $A$?
References


