ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

PROJECTIVE NOISE REDUCTION WITH DYNAMIC NEIGHBORHOOD SELECTION

A. Kern, D. Blank, and R. Stoop

Institut für Neuroninformatik,
University of Zurich / ETH Zurich,
CH-8057 Zürich
(albert|dan|ruedi)@ini.phys.ethz.ch

ABSTRACT

In recent years, several methods of noise reduction have been devised and applied to chaotic time series. Among them, projective methods have been particularly effective. We explain the non-orthogonal projective approach originally suggested by Grassberger et al. from a new point of view. We further discuss the extent to which a dynamic neighborhood selection improves noise reduction results.

1. INTRODUCTION

Chaotic signals are often contaminated with noise, which may either originate from measurement errors or from secondary parameters, whose influence on the system is small and should be neglected. The presence of noise limits our ability to extract information from the system. In particular, the performance of chaotic communication schemes [1] is reduced if the signal is noisy. Algorithms that are capable to reduce the noise in time series have therefore gained increased interest in recent years.

Among these algorithms generally known as "noise cleaning methods", local projective schemes have proven very efficient. The first step in nonlinear noise reduction is the reconstruction of the higher dimensional dynamics by means of the coordinate delay method [2] from a time series $(x_i)_{i=1}^N$. The reconstruction is obtained by defining m-dimensional vectors $\mathbf{x_i} = (x_i, x_{i-T}, \ldots, x_{i-(m-1)T})$ with suitable delays T, which is called the embedding process. The points in the embedding space of dimension m, give a diffeomorphic representation of the original, unknown, system, if $m >= 2D + 1$, where D is the box-counting dimension (Whitney's embedding theorem, [3]).

The common principle of all local projective methods is the following: Locally, optimal linear approximations to the invariant unstable manifolds of the unperturbed dynamics are evaluated in the embedding space, by using locally available noise-contaminated embedded data. After the local structures are known, the points are projected on the linear subspaces that locally span the attractor submanifold. The crucial issue that distinguishes different methods is with regard to what criterion the linear approximation is chosen to be optimal. This choice entails unexpected consequences and results in essentially two different ways of how the projections are actually performed: orthogonal and non-orthogonal. In this paper, we first present evidence for the assumption that orthogonal projections may not be optimal. Then we explain the non-orthogonal projective scheme suggested by Grassberger et al. [4] in Sec. 3 from a new, projection, point of view. Finally, we present results from a modification of the original algorithm and show that a measurable improvement for moderate noise levels can be obtained.

2. ORTHOGONAL PROJECTIONS ARE NON-OPTIMAL

The most straightforward approach for the computation of the approximating subspaces, would naturally be a local SVD (singular value decomposition) [5]. This scheme is optimal with respect to minimized least squares deviations in the embedding space and implies that orthogonal projections are performed. However, recently Schimming and Hasler [6] were able to analytically show that for piecewise linear one-dimensional maps, orthogonal projections are not optimal. In their case, they had complete knowledge of the unperturbed dynamics, and they used the minimum Bayesian variance estimator correction as the relevant criterion. They showed that the correction can be separated into two contributions: 1) An orthogonal projection to the linear attracting manifold, and 2) a so-called boundary function. This second contribution is important near turning points of the manifold and leads to the fact that the projections are non-orthogonal (Fig. 1, points B and C). Therefore, for non-linear maps, the a-priori assumption of orthogonal projections must be given up.

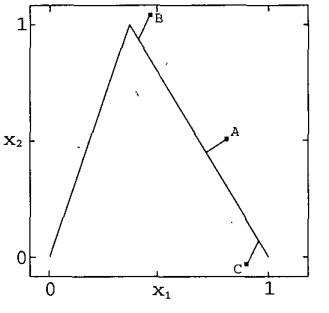

Figure 1: Noise reduction by minimum Bayesian variance estimation on the skew tent-map. Adapted from Schimming and Hasler [6].

Grassberger's approach, however, was originally motivated from another observation. The influence of noise is not equal for all components of an embedded data point. Consider the coordinate difference between the point and its next neighbor as the signal that is to be compared with the noise. Typically, close neighbors will

0-7803-5482-6/99/$10.00 ©2000 IEEE

have very close central coordinates, while the more peripheral coordinates will be more different (all the points are unstable saddle points). Therefore, the component with the highest noise-to-signal ratio is the central one (or the most central ones). As a consequence, noise reduction should concentrate on the central coordinates, while peripheral coordinates should be spared from noise reduction. Clearly, also applying noise reduction in a coordinate-specific way makes the associated projections non-orthogonal.

3. OPTIMAL CORRECTIONS BY NON-ORTHOGONAL PROJECTIONS

The core idea of the non-orthogonal projective approach is to find local linear subspaces of dimension $m - Q$ to approximate the unperturbed dynamics. For the algorithm, Q has to be chosen consistently with the other parameters (see Sec. 4), and the corrections $\Delta\mathbf{x}_n$ are chosen to place the points $\mathbf{x}'_n = \mathbf{x}_n - \Delta\mathbf{x}_n$ on the estimated manifolds. This can be achieved by imposing the Q constraints on $\Delta\mathbf{x}_n$,

$$\mathbf{b}_q^T \cdot (\mathbf{w}_n - \Delta\mathbf{x}_n) = 0, \qquad q = 1, .., Q, \tag{1}$$

where $\mathbf{w}_n = \mathbf{x}_n - \bar{\mathbf{x}}^{(n)}$ and $\bar{\mathbf{x}}^{(n)}$ denotes the center of mass of a neighborhood $\mathcal{U}_n$ around $\mathbf{x}_n$ of optimal size. The main problem is to choose the parameters and to estimate the vectors $\mathbf{b}_q$ and the corrections $\Delta\mathbf{x}_n$ in a self-consistent way. If the vectors $\mathbf{b}_q$ were equal to the rows of the Jacobian matrix of the dynamics at $\bar{\mathbf{x}}^{(n)}$, the corrections would be compatible with the time evolution. However, since the dynamics of the system is normally not known, this is generally not the case.

Component-specific corrections may be obtained by a transformation of the problem (1) by a diagonal matrix $\mathbf{R}$ (of dimension m), whose elements will be determined according to the weights we intend to give to the coordinates,

$$(\mathbf{z}_n - \Delta\mathbf{z}_n) := \mathbf{R}(\mathbf{w}_n - \Delta\mathbf{x}_n). \tag{2}$$

If matrix $\mathbf{R}$ were chosen as the identity matrix $\mathbf{I}$, we would remain with the previous situation. If $\mathbf{R}$ had only zero and unit entries on the diagonal, $\mathbf{R}$ would be a projection. The final choice will be that of a compromise between the two situations, where we $\mathbf{R}$ will be kept invertible, but where the entries are either large or small, which makes $\mathbf{R}$ act essentially like a projection. In the transformed space we require, similarly to (1), the fundamental relation

$$\mathbf{a}_q^T \cdot (\mathbf{z}_n - \Delta\mathbf{z}_n) = 0, \qquad q = 1, .., Q, \tag{3}$$

where the $\mathbf{a}_q$'s are chosen to be orthonormalized. If the directions $\mathbf{a}_q$ are known, the corrections $\Delta\mathbf{z}_n$ are given by the component of $\mathbf{z}_n$ lying in the spanned subspace:

$$\Delta\mathbf{z}_n = \sum_{q=1}^{Q} \mathbf{a}_q(\mathbf{a}_q^T \cdot \mathbf{z}_n). \tag{4}$$

When the condition (3) in the $\mathbf{z}$-space is expressed in terms of the $\mathbf{x}$-space, this amounts to

$$\mathbf{a}_q^T \cdot \mathbf{R}(\mathbf{w}_n - \Delta\mathbf{x}_n) = 0, \qquad q = 1, .., Q, \tag{5}$$

which with $\mathbf{a}_q =: \mathbf{R}^{-1}\mathbf{b}_q$ can be written as

$$\mathbf{R}^{-1}\mathbf{b}_q^T \cdot \mathbf{R}(\mathbf{x}_n - \Delta\mathbf{x}_n) = 0, \qquad q = 1, .., Q, \tag{6}$$

from which we finally obtain (1). Note, that the vectors $\mathbf{b}_q$ and the corrections $\Delta\mathbf{x}_n$ now depend on the choice of the matrix $\mathbf{R}$. If $\mathbf{R} = \mathbf{I}$, the set $\{\mathbf{b}_q\}_{q=1}^{Q}$ would be orthogonal, and, by (4), orthogonal corrections are obtained. Generally, however, we obtain non-orthogonal corrections. For an optimal correction in the $\mathbf{z}$-space,

$$\sum_{k:\,\mathbf{x}_k \in \mathcal{U}_n} (\sum_{q=1}^{Q} \mathbf{a}_q(\mathbf{a}_q^T \cdot \mathbf{z}_k))^2 = \min. \tag{7}$$

must be solved, under the condition

$$\mathbf{a}_q^T \cdot \mathbf{a}_{q'} = \delta_{qq'}. \tag{8}$$

The solution of this minimization problem is given by the eigenvectors of the covariance matrix

$$\Gamma_{ij} = \frac{1}{|\mathcal{U}_n|} \sum_{k:\,\mathbf{x}_k \in \mathcal{U}_n} (\mathbf{z}_k)_i (\mathbf{z}_k)_j, \tag{9}$$

with the Q smallest eigenvalues. They identify the directions $\mathbf{a}_q$ that satisfy Eq. (3), where the subspace spanned by them coincides with the nullspace of the operator Γ for the unperturbed system. Using Eq. (4), the corrected points are finally given in terms of the original vector $\mathbf{x}_n$ as

$$\mathbf{x'}_n = \mathbf{x}_n - \mathbf{R}^{-1} \sum_{q=1}^{Q} \mathbf{a}_q[\mathbf{a}_q^T \cdot \mathbf{R}(\mathbf{w}_n)]. \tag{10}$$

Closer examination of (10) reveals the basic structure of the optimized noise reduction procedure: First, a linear transform $\mathbf{z} = \mathbf{R}\mathbf{x}$ is performed. Then orthogonal projections are performed in the $\mathbf{z}$-space. The equation

$$\mathbf{b}_q^T \cdot \mathbf{P}\mathbf{b}_{q'} = \delta_{qq'}, \tag{11}$$

then immediately follows by expressing the orthonormalization of the set $\mathbf{a}_q$ in terms of $\mathbf{b}_q$, where we used $\mathbf{P} := \mathbf{R}^{-2}$. It shows that the set $\{\mathbf{b}_q\}_{q=1}^{Q}$ is orthonormalized in the metric defined by $\mathbf{P}$ rather than in the Euclidean metric, for $\mathbf{P} \neq \mathbf{I}$. Finally, by multiplication with $\mathbf{R}^{-1}$, the result is transformed back into the $\mathbf{x}$-space. If the peripheral entries of $\mathbf{R}$ are chosen as large and the central entries as small numbers, $\mathbf{R}$ can be inverted and $\mathbf{R}^{-1}$ would approximately project the corrections on the manifold spanned by the central coordinates in the $\mathbf{x}$-space, as required earlier.

However, since the corrections will generally not be compatible with the time evolution, further iterations of this procedure are required. The hope is to achieve convergence towards zero corrections by the iterative application of the corrections.

4. OPTIMAL CHOICES OF R AND Q

What are the implications imposed on the transformation $\mathbf{R}$? Certainly, Eq. (11) should be fulfilled. Assuming that the vectors $\mathbf{b}_q$ equal the rows of the Jacobian, in view of the normalization condition (11) the role of $\mathbf{R}$ (or, more precisely that of $\mathbf{R}^{-1}$) would then be to compensate for the action of the Jacobian on the Q coordinates. Unfortunately, we have no access to the Jacobian matrix. However, we may calculate the set of $\mathbf{a}_q$'s, once we have chosen $\mathbf{R}$ and Q. Fortunately, the minimization problem (7) ensures that

the noise-cleaned trajectory is optimal in the least-squares sense in the $\mathbf{z}$-space, for every chosen $\mathbf{R}$. To obtain optimally weighted non-orthogonal noise cleaning, $\mathbf{R}$ and Q, are chosen along the following rules:

If we intend to suppress d peripheral coordinates of the correction vector $\Delta\mathbf{x}_j$, where $j = 1, .., m$ indexes the components of the vector, we choose the corresponding elements R_{jj} very large in comparison to the central ones. In practice, it is good enough to choose these elements a factor of $10 - 1000$ times larger than the remaining ones. In view of making projections onto the approximating local linear subspaces of dimension $m - Q$, the number of constraints Q that we can impose is limited: The vector $\Delta\mathbf{x}$ that establishes the projection is of dimension m, although only $m - d$ of its components may be changed. If no components are left to be changed (i.e., $d = m$), then the projection fails. By the projection, the Q constraints should be met. This clearly is only possible if at least Q components can be changed. Since we may change $m - d$ components, the relation $Q \leq m - d$ needs to be satisfied. Therefore, the relevant conditions for the choice of $\mathbf{R}$ and Q are

$$\begin{aligned} Q &\leq m - m_0, \\ Q &\leq m - d, \end{aligned} \tag{12}$$

where, in order to leave no uncorrected noisy directions, the optimal choice is the largest possible Q:

$$Q = m - \max\{m_0, d\}. \tag{13}$$

Below we give a practical illustration (see Fig. 2). For given $m = 5$, we need to estimate m_0 and set d (which may be estimated from the correlation dimension, and from the Lyapunov exponents (compared to the relative size of the entries of $\mathbf{R}$), respectively, if available). Then Q is determined according to the above relation. For Hénon's data, $m_0 = 2$, $d = 2$, and $R_{jj} = 1$ for $j = 2, 3, 4$ and $R_{jj} = 100$, for $j = 1, 5$. Other choices that are compatible with (12) would be $d = 0$ (implying orthogonal projections) and $d = 4$, that are both non-optimal.

5. TRAJECTORY RECONSTRUCTION

Still under the assumption that the dynamics does not dominate the noise over d time steps, where d is the number of the to be corrected coordinates, the different corrections that are obtained for the same time series elements x_i, display an approximate random distribution. In the non-orthogonal projective algorithm, by means of $\mathbf{R}$ different corrections are applied to the components. Since the corrections with the largest weights are the most reliable ones, this should be accounted for when taking the average, and weighted averages should be taken. Remembering that large entries in $\mathbf{R}$ lead to the suppression of the corresponding component corrections, the weights

$$w_i = \frac{\mathrm{Tr}\,\mathbf{R}}{R_{ii}} \tag{14}$$

are the appropriate choice for making the average for the corrected i'th time series element. Explicitly, the i'th time series element correction is chosen as

$$\Delta x_i{}^{TS} = \sum_{r=1}^{m} w_r \,\mathrm{sel}[\Delta\mathbf{x}_{i,r}, r], \tag{15}$$

where $\Delta x_{i,r}$ are the corrections of the corresponding coordinate when appearing in the m points of the embedding space as coordinate r, respectively and sel$[\mathbf{u}, k]$ is the function that selects the i'th component from vector u.

In the first pass of the noise reduction algorithm, the average is performed over considerably different corrections sel$[\Delta\mathbf{x}_{i,r}, r]$, $r = 1, .., m$ to the same time series element. Further iterations of the procedure ensure that the different representatives of the same time series element converge to one value. For Hénon's data, we found that 5 iterations are sufficient to converge, for all considered noise levels. In the different iterations, the values Q, d and $\mathbf{R}$ were kept fixed. As an example, the effect of one iteration on the Hénon system, to which noise with 0.01 standard deviation was added, is shown in Fig. 2.

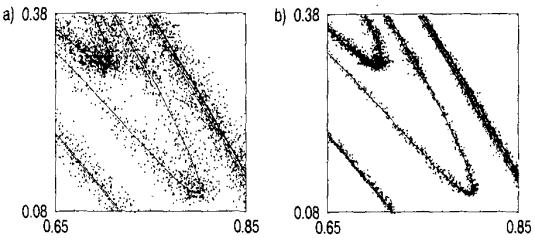

Figure 2: Results of non-orthogonal noise reduction, illustrated by means of data from the Hénon attractor with additive Gaussian noise of $\sigma = 0.01$ std. dev. The unperturbed attractor structure are the superimposed lines. a) Original data. b) Results after one iteration.

6. IMPLEMENTATION

In the first step of the algorithm, a neighborhood around each data point $\mathbf{x}_n$ is required that contains at least N_{min} points; the size of the neighborhood should be at most ϵ_{max} (to guarantee the linear regime) and should contain at least $2m + 1$ data points to ensure good statistics for calculating the covariance matrix Γ. This can be achieved with an efficient neighbor search algorithm. In the next step, the covariance matrix (9) and its eigenvectors and eigenvalues are computed, by using standard tools from linear algebra. Next, the corrections (10) are calculated. In order to obtain good statistics, N_{min} and ϵ_{min} should not be too small. We found that $N_{min} = 50$ is a good choice for the first iteration. For subsequent iterations, when the points are pushed together by noise reduction, smaller values for N_{min} should be chosen. Fortunately, the algorithm is insensitive to small variations in the values for N_{min}, ϵ_{min}, and ϵ_{max}. Reasonable results can be obtained for whole ranges of these parameters.

As discussed above, a crucial issue in the implementation of the non-orthogonal projective noise reduction is the selection of appropriate neighborhoods. Although the algorithm is rather stable with respect to the parameters defining the neighborhood size, we propose a sophistication of the neighborhood selection method, to achieve a further improvement of the already excellent algorithm. In particular, we propose a dynamic neighborhood selection that assigns to each point a neighborhood of a locally optimized size.

There are several ways how this can be done in an efficient way. Starting from a neighborhood $\mathcal{U}_n$ around a point $\mathbf{x}_n$, we need to find a reliable measure for the quality of the approximation of the (unperturbed) linearized dynamics (in the $\mathbf{x}$-, $\mathbf{z}$-spaces), by the linear subspaces $E_{\mathbf{x}_n}$, $E_{\mathbf{z}_n}$ that are evaluated from the noisy data.

$E_{\mathbf{z}_n}$ is well defined if the eigenvalues of the eigenvectors of Γ extending into the "noisy" directions are much smaller than its largest eigenvalue. However, if $\mathcal{U}_n$ is too large, $E_{\mathbf{z}_n}$ only provides a poor approximation to the local linearized manifold due to the strong influence of the nonlinearities for large neighborhoods. On the other hand, the approximation is bad if all eigenvalues are of comparable magnitudes. This will happen if $\mathcal{U}_n$ contains an insufficient number of points, or if it extends only over a range that is comparable to the noise level. In this case, the linear space $E_{\mathbf{z}_n}$ approximating the dynamics cannot properly be determined. For this reason, the ratio of the largest "noisy" eigenvalue, λ_{noise}, to the largest eigenvalue pertaining to the principal dynamics, λ_{max}, as a function of the neighborhood size ϵ can serve as a measure for the quality of the linear approximation, and local minima of $\lambda_{noise}(\epsilon)/\lambda_{max}(\epsilon)$ are candidates for an optimal neighborhood size, ϵ_{opt}. In our algorithm, ϵ_{opt} is determined automatically for each point, by optimizing $\lambda_{noise}(\epsilon)/\lambda_{max}(\epsilon)$, while the selection of too large values for ϵ_{opt} is prohibited.

7. RESULTS

We tested the non-orthogonal projective algorithm with dynamic neighborhood selection on the Hénon data, for additive Gaussian noise with standard deviations between 0.005 and 0.02. In all cases, the central $m - d = 3$ coordinates were corrected and $d = 2$ coordinates remained uncorrected. According to (13), for the number of constraints $Q = 3$ emerges. For noise with standard deviation $\in [0.07, 0.013]$, our algorithm with dynamic neighborhood selection yielded noticeably better noise reduction if compared to the non-dynamic variant. The meandering of the results around the true attractor structures, which is observed for both algorithms, is significantly suppressed by the dynamic approach, and the corrected points generally lie closer to the ideal manifold. Furthermore, the blurring in the structure, that is sometimes observed, is significantly reduced by the dynamic method. Also, in the vicinity of tangency points our algorithm shows an improved performance. In Fig. 3 we give two illustrations for the performance of the noise reduction with dynamic neighborhood selection. It can be seen that fine details of the original attractor structure can be recovered.

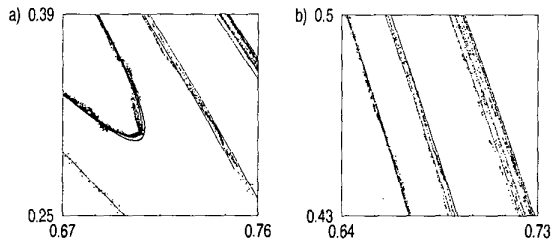

Figure 3: Effects of the dynamic noise reduction method on two characteristic sections of the Hénon attractor after 4 iterations (additive Gaussian noise with $\sigma = 0.01$ std. dev).

However, the advantages of the neighborhood selection sophistication are restricted to noise from a relatively small interval of standard deviations. For levels of noise beyond 0.013, the noisy attractor shows almost no fine structures, and the dynamic neighborhood selection is therefore of no advantage. If the noise level is smaller than 0.07, a relatively small number of neighboring points already yields good statistics for the approximation of the true attractor structure. In most cases, only points belonging to the to $\mathbf{x}_n$ closest piece of the expanding manifold will be recruited. In this situation, the methods will produce practically identical results.

In conclusion, a considerable improvement at moderate noise levels is observed. For other noise levels the use of the dynamic neighborhood selection is not justified and only results in an increased amount of calculation time and computational resources. For large noise levels, we recommend as the most sophisticated approach pre-filtering in combination with the dynamic neighborhood selection algorithm.

8. ACKNOWLEDGEMENTS

We acknowledge support from the Swiss National Science Foundation (KTI-contract with Phonak AG).

9. REFERENCES

[1] L.M. Pecora and T.L. Carroll, Phys. Rev. Lett. **64**, 821 (1990);
C.W. Wu and L.O. Chua, Int. J. of Bif. and Chaos **3**, 1619 (1993).

[2] J. Peinke, J. Parisi, O.E. Roessler, and R. Stoop, *Encounter with chaos*, Springer, Berlin (1992).

[3] H. Whitney, Ann. Math. **37**, 645 (1936).

[4] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath, and T. Schreiber, Chaos **3** (2), 127 (1993) (general non-orthogonal projections);
H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge (1997) (implementation of the algorithm);
T. Schreiber and P. Grassberger, Phys. Lett. A **160**, 411 (1991) (noise reduction of one central coordinate).

[5] R. Cawley and G.-H. Hsu, Phys. Rev. A **46**, 3057 (1992);
R. Cawley and G.-H. Hsu, Phys. Lett. A **166**, 188 (1992);
T. Sauer, Physica D **58**, 193 (1992);
R. Stoop and J. Parisi, Physica D **50**, 89 (1991).

[6] Th. Schimming and M. Hasler, Proc. 7th Int. Workshop on Nonlinear Dynamics of Electronic Systems, Rønne, Denmark, 179 (1999).

[7] M. Hénon, Comm. Math. Phys. **50**, 69 (1976).