2-15-2005

Generalized Phase Space Projection for Nonlinear Noise Reduction

Michael T. Johnson
Marquette University, michael.johnson@marquette.edu

Richard J. Povinelli
Marquette University, richard.povinelli@marquette.edu

Physica D : Nonlinear Phenomena, Vol. 201, No. 3-4 (February 15, 2005): 306-317. DOI.

Generalized Phase Space Projection for Nonlinear Noise Reduction

Michael T. Johnson
Marquette University, EECE, Milwaukee, WI
Richard J. Povinelli
Marquette University, EECE, Milwaukee, WI

Abstract

Improved phase space projection methods, adapted from related work in the linear signal processing field based on subspace decomposition, are presented for application to the problem of additive noise reduction in the context of phase space analysis. These methods improve upon existing methods such as Broomhead–King singular spectrum analysis projection by minimizing overall signal distortion subject to constraints on the residual error, rather than using a direct least-squares fit. This results in a range of weighted projections which estimate and compensate for the portion of the principal component's singular values corresponding to noise rather than signal energy, and which include least-squares (LS) and least minimum mean square error (LMMSE) as subcases. The nature of phase space covariance, the key element in construction of the projection matrix, is

examined across global phase spaces as well as within local neighborhood regions. The resulting algorithm, illustrated on a noisy Henon map as well as on the task of speech enhancement, is applicable to a wide variety of nonlinear noise reduction tasks.

Keywords

Noise reduction, Singular spectrum analysis, Subspace decomposition

1. Introduction

The task of noise reduction is a central theme in a wide variety of fields. Methods for optimal signal/noise separation include linear signal processing techniques such as Wiener filtering and Kalman filtering as well as nonlinear methods such as manifold decomposition and phase space projection. There is often a great deal of overlap in the underlying concepts on which methods from separate areas are based, and significant improvements in both understanding and technique may be gained through cross-disciplinary study. In the work presented here, we establish a relationship between phase space projection methods commonly used on nonlinear dynamical systems and subspace decomposition methods commonly used in linear signal processing. From this relationship, we show that phase space techniques such as Broomhead–King projection can be improved by weighting the projection operator using a constraint-based minimization process.

A reconstructed phase space (RPS) matrix **Y** of dimension d and lag τ is called a trajectory matrix and defined by:

$$(1)\ \mathrm{Y} = \begin{bmatrix} \mathrm{y}_{1+(d-1)\tau} \\ \mathrm{y}_{2+(d-1)\tau} \\ \vdots \\ \mathrm{y}_{N} \end{bmatrix} = \begin{bmatrix} y_{1+(d-1)\tau} & \cdots & y_{1+\tau} & y_1 \\ y_{2+(d-1)\tau} & \cdots & y_{2+\tau} & y_2 \\ \vdots & & \ddots & \\ y_N & \cdots & y_{N-(d-2)\tau} & y_{N-(d-1)\tau} \end{bmatrix}_{(N-(d-1)\tau)\times d},$$

where the row vectors $\mathbf{y}_n$, with $n = (1 + (d - 1)\tau) \ldots N$, represent individual points in the RPS:

$$(2)\ \mathrm{y}_n = [y_{n-(d-1)\tau} \quad \cdots \quad y_{n-\tau} \quad y_n].$$

The fundamental nature of an RPS is captured by a characteristic pattern traced out in the space, the orbit or trajectory of the system. In dissipative systems, there exists a bounded subset of the phase space to which trajectories asymptote as time increases, called an attractor [1]. The motivation to use RPSs for analysis lies with historical work in topology and nonlinear systems analysis [2], [3], [4], [5], [6], which has shown that when constructed properly RPSs are homeomorphic embeddings of the underlying system's state space. RPSs may be constructed from either a single or multiple time-series, and there are a variety of methods available to estimate the appropriate lag and dimension from data [7], [8]. RPSs can be a useful tool in the analysis of many kinds of systems, especially those with nonlinearities that traditional frequency domain analysis cannot capture. For an introduction to the analysis and behavior of nonlinear dynamical systems in general and chaotic systems in particular see [1], [9], [10]. For an overview of analysis and applications using reconstructed phase spaces see [7], [11], [12].

2. Noise models and review of nonlinear noise reduction techniques

Noise reduction is a key area where RPS analysis can be a powerful tool. An observed time series signal with additive noise is given by

$$(3)\ \mathrm{y} = \mathrm{x} + \mathrm{w},$$

where $\mathrm{y} \triangleq \{y_n, n = 1 \ldots N\}$ represents the observed signal, and the unknown clean signal **x** and additive noise **w** are assumed to be independent. This directly equates to the trajectory matrix relationship

(4) $Y = X + W$,

where **Y**, **X** and **W** are the corresponding time-delay RPSs of each signal. It is assumed that **x** is itself generated by an underlying dynamical system given by

(5) $x_n = F(x_{n-1}) + v_{n-1}$,

where $\mathbf{x}_n$ is a point in the reconstructed phase space (or true phase space) representing the dynamical system and $\mathbf{v}_n$ an additive noise vector referred to as a "dynamical noise" component, since it is coupled into the system at each time step. In the absence of either dynamical noise or additive noise components, the observed time series is completely deterministic and is an exact representation of the underlying system **F** (which for nonlinear/chaotic systems may still have exhibit extremely complex behavior and a broad-band spectral composition).

The goal of noise reduction is essentially to find the best possible estimate of the noise-free signal. Current methods for accomplishing this task with nonlinear systems are primarily based on either projections of the phase space or on manifold decomposition along stable and unstable directions of the attractor.

In global Principal Components Analysis, also known as Singular Systems Analysis per Broomhead and King [13], a singular value decomposition $\mathbf{Y} = \mathbf{U\Sigma C}^H$ is taken on the trajectory matrix, where **C** is the eigenvector matrix of the trajectory covariance. A projected trajectory matrix is computed via the equation:

(6) $\hat{X} = YC_1C_1^H$,

where $\mathbf{C}_1$ consists of the columns of **C** such that the corresponding singular values are greater than the noise level threshold σw. The result is that the original attractor is projected onto the principal eigenvectors of the space. To implement this approach for noise reduction, the original time series is over-embedded, i.e. embedded into a dimension well over that required for attractor representation. The principal axes are determined with an SVD of the trajectory matrix, and a projection is done using Eq. (6) above. An enhanced one-dimensional signal is created from the new space, typically by time-aligning and averaging the columns of the trajectory matrix.

The concept of projection within the reconstructed phase space can be easily adapted to apply to local neighborhoods within the space. Schreiber and Grassberger [14], Cawley and Hsu [15], and Sauer [16] have all developed methods based on this idea, among which the primary differences are the mechanisms by which the dynamics of the system are approximated within the local neighborhood regions. For these approaches, the time series is over-embedded as with the global method, and then each point in the space is individually transformed using a projection based only on its local neighborhood region.

Another approach is noise reduction based on shadowing, introduced by Hammel [17] and extended by Farmer and Sidorowich [18], which is based on an a priori knowledge of the system function $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$. Given this knowledge of the system, the shadowing lemma [19] is used to identify a noise-free trajectory that is close to the original noisy trajectory by factoring the dynamics into stable (exponentially contracting) and unstable (exponentially expanding) directions, a process known as manifold decomposition. Provided that the system **F** is everywhere hyperbolic, this decomposition is always possible; however, if homoclinic tangencies or near-tangencies are present such that the angle between the stable and unstable directions is close to zero, the algorithm performs poorly and must be augmented by more robust and computationally expensive methods.

3. Linear subspace models

Subspace decomposition is a well-established technique in linear signal processing, based on the underlying assumption that a noise-free signal segment **x** of length *K* can be represented as a superposition of *M* linearly independent basis vectors

(7) $\mathrm{x} = \mathrm{Vs}$,

where **V** is a $K \times M$ matrix whose columns are basis vectors, and **s** the vector of coefficients. A typical set of bases for **V** might be the damped complex sinusoid model, with columns:

(8) $\mathrm{V}_m = [1 \quad \rho_m^1 \mathrm{e}^{j1\omega_m} \quad \cdots \quad \rho_m^{K-1} \mathrm{e}^{j(K-1)\omega_m}]^{\mathrm{T}}$,

where $0 \le \rho_m \le 1$ is a damping coefficient and $0 \le \omega_m \le 2\pi$ the mth basis frequency. The additive noise model of (3) can be analyzed through the use of the resulting autocorrelation function equation:

(9) $\mathrm{R_y} = \mathrm{R_x} + \mathrm{R_w}$,

where $\mathbf{R_y} = E\{\mathbf{y}\mathbf{y}^H\}$, and $\mathbf{R_x}$ and $\mathbf{R_w}$ are defined similarly, each $K \times K$ matrices. The covariance matrix $\mathbf{R_y}$ is a Toeplitz matrix of autocorrelation values

(10) $$\mathrm{R_y} = \begin{bmatrix} r(0) & r(1) & \cdots & r(K-1) \\ r(-1) & r(0) & & r(K-2) \\ \vdots & & \ddots & \\ r(-K+1) & r(-K+2) & & r(0) \end{bmatrix},$$

which can be estimated using a temporal average from the windowed data matrix

(11) $$\mathrm{A}^H = \begin{bmatrix} y_K & y_{K+1} & \cdots & y_N \\ y_{K-1} & y_K & & y_{N-1} \\ \vdots & & \ddots & \\ y_1 & y_2 & & y_{N-K+1} \end{bmatrix}_{K \times (N-K+1)},$$

through the relationship

(12) $\mathrm{R_y} \approx \frac{1}{N-K+1} \Phi = \frac{1}{N-K+1} \mathrm{A^H A}$,

where $\Phi \triangleq \mathrm{A}^H\mathrm{A}$ is known as the correlation matrix. Reviewing Eq. (1), we see that the data matrix **A** and trajectory matrix **Y** have the same structure, so that for lag $\tau = 1$ and dimension $d = K$, we have **A** = **Y** and $\mathbf{R_y}$ = **C**, the trajectory covariance matrix.

Without loss of generality, we assume that the noise is white, since if this is not the case it can be whitened by applying $\mathrm{R_w}^{-1/2}$, known a priori or estimated from segments without signal content, to the system [20]. Given this, the covariance of the noise is simply $\mathrm{R_w} = \sigma_w^2 \mathrm{I}_K$, where σ_w^2 is the additive noise power. Another consequence of this is that the eigenvectors of $\mathbf{R_y}$ are also eigenvectors of $\mathbf{R_x}$ and $\mathbf{R_w}$. For real time series data, these covariance matrices will always be positive definite and Toeplitz. Performing a Karhunen–Loeve Transform (KLT) on $\mathbf{R_y}$, we see that the decomposition of the noisy covariance matrix leads to

(13) $\mathrm{R_y} = \mathrm{R_x} + \mathrm{R_w} = \mathrm{Q}\Lambda_x\mathrm{Q}^H + \sigma_w^2\mathrm{Q}\mathrm{I}_K\mathrm{Q}^H = \mathrm{Q}(\Lambda_x + \sigma_w^2\mathrm{I}_K)\mathrm{Q}^H$.

Given (7) and the assumption of no zero-valued coefficients in **s**, the rank of $\mathbf{R_x}$ will be *M*, so there will always be *M* positive eigenvalues and $(K - M)$ zero eigenvalues. This leads to a KLT that can be written in the block matrix form:

(14) $\mathrm{R_y} = [\mathrm{Q_1} \quad \mathrm{Q_2}] \begin{bmatrix} (\Lambda_{x_1} + \sigma_w^2 \mathrm{I}_M) & 0 \\ 0 & \sigma_w^2 \mathrm{I}_{(K-M)} \end{bmatrix} \begin{bmatrix} \mathrm{Q}_1^H \\ \mathrm{Q}_2^H \end{bmatrix}$.

If the dimension *M* is known a priori it is possible to estimate $\mathbf{R_x}$ through the equation:

(15) $\mathrm{R_x} = \mathrm{Q_1}(\Lambda_{\mathrm{y}_1} - \sigma_w^2 \mathrm{I}_M)\mathrm{Q}_1^H$,

which, assuming that the eigenvalues are ordered in a non-increasing manner through a robust decomposition method such as a singular value decomposition (SVD), also ensures that the computed eigenvalues are positive and the covariance matrix remains positive definite. Due to the underlying assumptions of this model, the space spanned by $\mathbf{Q}_1$ (and also therefore the space spanned by the signal basis matrix **V**) is called the *signal subspace*, and the complementary space spanned by $\mathbf{Q}_2$ is called the *noise subspace*, although the noise itself spans the entire space.

This analysis can be used in several ways, including for spectral estimation methods such as MUSIC and ESPRIT [21] as well as for signal enhancement [22]. To perform enhancement, we can for example directly project the noisy signal onto the signal subspace through $\hat{\mathrm{x}} = \mathrm{Q_1}\mathrm{Q}_1^H\mathrm{y}$, which is equivalent to a least-squares error fit. For an arbitrary linear projector:

(16) $\hat{\mathrm{x}} = \mathrm{Hy}$,

the error signal is given by:

(17) $e = \hat{\mathrm{x}} - \mathrm{x} = \mathrm{H}(\mathrm{x} + \mathrm{w}) - \mathrm{x} = (\mathrm{H} - \mathrm{I})\mathrm{x} + \mathrm{Hw} \triangleq e_x + e_w$,

where e_x is error due to signal distortion and ew the error due to residual noise. Ephraim and van Trees [22] derived several linear estimators to allow constraints on the type of error. In each case the form of the estimator is that of $\mathrm{H} = \mathrm{Q_1}\mathrm{G}\mathrm{Q}_1^H$, where **G** is a diagonal matrix of modified eigenvalues called the weighting matrix. We will focus on the time domain constrained (TDC) estimator given by:

(18) $\mathrm{H_{TDC}}: \min_H e_x^2 \text{ s.t. } \frac{1}{K} e_w^2 < \alpha\sigma_x^2, \mathrm{H_{TDC}} = \mathrm{Q_1} \frac{\Lambda_{x_1}}{\Lambda_{x_1} + \mu\sigma_w^2 \mathrm{I}_M} \mathrm{Q}_1^H = \mathrm{Q_1} \frac{\Lambda_{y_1} - \sigma_w^2 \mathrm{I}_M}{\Lambda_{\mathrm{y}_1} + (\mu-1)\sigma_w^2 \mathrm{I}_M} \mathrm{Q}_1^H$,

where $0 \le \alpha \le 1$ is a parameter limiting the residual noise and μ is the resulting transform parameter, which can be determined analytically as a Lagrange multiplier or set arbitrarily.

The well-known linear minimum mean-squared-error (LMMSE) and least-squares (LS) estimators are subcases of $\mathbf{H}_{\mathrm{TDC}}$. The LMMSE estimator [23] is given by $\mu = 1$:

(19) $\mathrm{H_{LMMSE}} = \mathrm{H_{TDC|}}_{\mu=1} = \mathrm{Q_1} \frac{\Lambda_{y_1} - \sigma_w^2 \mathrm{I}_M}{\Lambda_{y_1}} \mathrm{Q}_1^H$,

and the simpler least-squares (LS) estimator [24] is given by $\mu = 0$:

(20) $\mathrm{H_{LS}} = \mathrm{H_{TDC|}}_{\mu=0} = \mathrm{Q_1}\mathrm{Q}_1^H$.

In general, $\mathbf{H}_{\mathrm{LMMSE}}$ is preferred over $\mathbf{H}_{\mathrm{LS}}$ because it more accurately weights the contributions of each eigenvector to the signal subspace.

Although the various linear projectors above are derived for application to linear signals and covariance matrices, we will see that they apply equally to projection on arbitrary phase spaces.

4. Generalization to arbitrary lag RPS projection

4.1. Covariance analysis of global and local phase spaces

We have already seen that the trajectory matrix **Y** and data matrix **A** are equivalent except for the difference due to lag value. For arbitrary lag τ, the trajectory covariance matrix $\mathrm{R}_{\mathrm{Y}}^{\tau}$ is the covariance of the time delay signal vector $\mathbf{y}_n$:

$$(21)\ \mathrm{R}_{\mathrm{Y}}^{\tau} = \begin{bmatrix} r(0) & r(\tau) & \cdots & r((d-1)\tau) \\ r(-\tau) & r(0) & & r((d-2)\tau) \\ \vdots & & \ddots & \\ r(-(d-1)\tau) & r(-(d-2)\tau) & & r(0) \end{bmatrix} \approx \frac{1}{N-(d-1)\tau}\mathrm{Y}^{\mathsf{T}}\mathrm{Y}.$$

For the case of additive noise uncorrelated with the signal, we have $\mathrm{R}_{\mathrm{Y}}^{\tau} = \mathrm{R}_{\mathrm{X}}^{\tau} + \mathrm{R}_{\mathrm{W}}^{\tau}$, just as in Eq. (9), and if the noise is white the matrix $\mathrm{R}_{\mathrm{W}}^{\tau} = \sigma_w^2 \mathrm{I}_d$ and the decomposition equations (13)–(15) hold as well. In general, the impact of dynamical noise is not purely additive with respect to the covariance matrix, but rather depends on the system **F** and its Jacobian within individual neighborhoods of the attractor. This impact is highly correlated and will therefore alter the eigen-decompositions of $\mathrm{R}_{\mathrm{Y}}^{\tau}$.

Within a local neighborhood of the embedding space, the original covariance $\mathrm{R}_{\mathrm{X}}^{\tau}$ and noisy covariance $\mathrm{R}_{\mathrm{Y}}^{\tau}$ are different from those of the global space, representing the covariance of a subset of the trajectory row vectors rather than that of a continuous time series. However, the additivity property of Eq. (9) still holds, so that given the same assumptions as for the global case the decomposition equations (13)–(15) are valid and can be used to derive projection operators in the same way, with the constraints applying to neighborhoods around individual points rather than the entire signal.

Overall, these covariance matrix properties hold for the general lag case both globally and within local neighborhoods. Thus selecting a noise reduction projection can be thought of as a tradeoff between signal distortion and residual error.

4.2. Projection operators

The optimal projections **H** with respect to the time delay RPS are the same as those for the subspace method, given in Eqs. (18), (19), (20). This results in $\hat{\mathrm{X}} = \mathrm{YH}_{\tau}$, where the subscript is used to indicate the lag of the trajectory matrix decomposition used to compute the projection matrix. Using the same minimization constraints as Ephraim and van Trees [22], the generalizations of Eqs. (18), (19), (20) for the case of arbitrary lag covariance matrices are given by:

$$(22)\ \mathrm{H}_{\mathrm{TDC},\tau} = \mathrm{Q}_1 \frac{\Lambda_{\mathrm{y}_1} - \sigma_w^2 \mathrm{I}_M}{\Lambda_{\mathrm{y}_1} + (\mu-1)\sigma_w^2 \mathrm{I}_M} \mathrm{Q}_1^H,$$

$$(23)\ \mathrm{H}_{\mathrm{LMMSE},\tau} = \mathrm{Q}_1 \frac{\Lambda_{y_1} - \sigma_w^2 \mathrm{I}_M}{\Lambda_{\mathrm{y}_1}} \mathrm{Q}_1^H,$$

$$(24)\ \mathrm{H}_{\mathrm{LS},\tau} = \mathrm{Q}_1 \mathrm{Q}_1^H.$$

We see by examination then that as expected the Broomhead–King singular spectrum projection given in Eq. (6) is equivalent in the lag one case to global $\mathbf{H}_{\mathrm{LS}}$ projection.

The above equations are valid for reconstructed phase spaces from any system. In the case of linear systems, if the rank of $\mathbf{R}_{\mathbf{y}}$ is known the identification of a model order *M* to partition the eigenvector and eigenvalue matrices is substantially simplified. There are several methods for determination of the model order if this rank is unknown, including approaches based on the minimum description length (MDL) principal [25], information theoretic metrics [26], or Bayesian estimation [27]. Similar methods, such as the false nearest neighbor

approach [7], exist for estimating the dimension of nonlinear systems, although these tend to be more heuristic than linear methods.

If the signal is over-embedded, such that the dimension of the RPS is significantly larger than the underlying dimension of the space, the eigenvalues will tend to asymptote to σw2, which allows for direct estimation of the noise power. The noise power can also often be estimated directly from regions containing noise but no signal.

5. Time-alignment and averaging to create univariate time-series

Once linearly transformed, the resulting trajectory matrix $\hat{X}$ no longer corresponds to a time-delay embedding on a single time series, and there is no unique mapping back to a one-dimensional signal. Each column of the transformed matrix represents a possible enhanced signal output. There are several methods for creating an enhanced time-series from the projected RPS, including selecting a single column from the new trajectory matrix or doing a time-aligned averaging of the columns.

In the work presented here, we use a time-aligned weighted average, with higher weight given to the values in the center columns of the matrix and lower weight given to the values in the left-most and right-most columns. This corresponds to emphasizing the time-centered value of each projected point. Signal points near the beginning or end of the trajectory matrix have fewer representatives and are weighted accordingly. To illustrate this process, an example trajectory matrix $\hat{X}$, aligned trajectory matrix $\hat{X}_{\text{aligned}}$, and weighting matrix **P** are shown below for the $d = 4$, $\tau = 2$ case with 100 points in the time-series.

$$\hat{X}^{T} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdots & 90 & 91 & 92 & 93 & 94 \\ 3 & 4 & 5 & 6 & 7 & 8 & \cdots & 92 & 93 & 94 & 95 & 96 \\ 5 & 6 & 7 & 8 & 9 & 10 & \cdots & 94 & 95 & 96 & 97 & 98 \\ 7 & 8 & 9 & 10 & 11 & 12 & \cdots & 96 & 97 & 98 & 99 & 100 \end{bmatrix}, \hat{X}^{T}_{\text{aligned}}$$

$$= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots & 94 & & & & & & \\ & & 3 & 4 & 5 & 6 & 7 & \cdots & 94 & 95 & 96 & & & & \\ & & & & 5 & 6 & 7 & \cdots & 94 & 95 & 96 & 97 & 98 & & \\ & & & & & & 7 & \cdots & 94 & 95 & 96 & 97 & 98 & 99 & 100 \end{bmatrix}, P^{T}$$

$$= \begin{bmatrix} 1 & 1 & .5 & .5 & .25 & .25 & .14 & \cdots & .14 & & & & & & \\ & & .5 & .5 & .5 & .5 & .36 & \cdots & .36 & .25 & .25 & & & & \\ & & & & .25 & .25 & .36 & \cdots & .36 & .5 & .5 & .5 & .5 & & \\ & & & & & & .14 & \cdots & .14 & .25 & .25 & .5 & .5 & 1 & 1 \end{bmatrix}.$$

The resulting output time-series is given by:

(25) $\hat{x} = \sum_{i=1}^{d} p_i \cdot \hat{x}_i$,

where $\hat{x}_i$ is the *i*th column from $\hat{X}_{\text{aligned}}$, $\mathbf{p}_i$ is the corresponding weight vector from **P**, and the $(\cdot\cdot)$ operator represents pointwise multiplication. The rows of the weighting matrix each sum to one.

There are several benefits to this approach, including reduced overall residual error resulting from averaging the error across multiple instances as well as the benefit of having an enhanced time series that exactly matches the original in total length.

As the time lag used in constructing the RPS increases, the benefits of weighted averaging decrease. This is caused by an increase in the time range between the centers of the rows being averaged, which causes fast moving dynamics to be averaged along with the residual noise and creates a source of additional error not originally present. For example, with dimension $d = 5$ and lag $\tau = 1$, the trajectory matrix rows which include the time point $\hat{x}_{100}$ range from $x_{100} = [\hat{x}_{96} \quad \hat{x}_{97} \quad \hat{x}_{98} \quad \hat{x}_{99} \quad \hat{x}_{100}]$ to $\hat{x}_{104} = [\hat{x}_{100} \quad \hat{x}_{101} \quad \hat{x}_{102} \quad \hat{x}_{103} \quad \hat{x}_{104}]$, with time-alignment windows extending over a five time point range. For the same dimension with lag $\tau = 20$,

the rows which include the time point $\hat{x}_{100}$ range from $\hat{\mathbf{x}}_{100} = [\hat{x}_{20} \quad \hat{x}_{40} \quad \hat{x}_{60} \quad \hat{x}_{80} \quad \hat{x}_{100}]$ to $\hat{\mathbf{x}}_{180} = [\hat{x}_{100} \quad \hat{x}_{120} \quad \hat{x}_{140} \quad \hat{x}_{160} \quad \hat{x}_{180}]$, with time-alignment windows extending over an 81 time point range. In general, the net alignment window of all matrix rows including a given time index has a range of $1 + (d - 1)\tau$. With the exception of artificially constructed systems designed to have good representation at specific lag values, a lag of $\tau = 1$ tends to give the best enhancement results, and this holds up empirically for both linear and nonlinear systems across weighting configurations. If the goal is phase space enhancement as opposed to time-series enhancement, weighted averaging is not needed and any lag can be used.

6. Nonstationarity issues

For stationary signals, signal enhancement is accomplished through direct application of the desired linear transformation, followed by weighted averaging of the resulting trajectory matrix as just outlined. Global phase space projections require only a single transform to the trajectory matrix, whereas local projections require identification of neighborhood regions and computation of transformation matrices for each point individually.

For signals with time-varying generating mechanisms, the situation is much more complex. Since the phase space attractor is changing structure over time, an embedding of the entire time series requires a much higher dimension, or else local neighborhood regions may include points from multiple temporal contexts, representing very different attractor patterns. It is often better to reduce attractor variation due to nonstationarity by defining the phase space, within the context of some predefined temporal or window region. To do this, the original time-series can be divided up into windows, each of which can be individually projected.

Since disjoint windows result in edge effects, the most common approach is to use overlapping analysis windows which can be recombined through an overlap-add method. In this process, each window is individually transformed using its own covariance estimate, projected to the desired dimension, and then re-combined with adjacent windows. Triangular windows and Hanning windows are both commonly used for this purpose, since when overlapped by exactly 50% these particular window types maintain a net additive weight of unity throughout.

A second stationarity issue that affects local projections is the mechanism for maintaining neighborhood region information. This process is computationally intense, but can be substantially improved by using efficient data structures and search algorithms [28]. For nonstationary signals, to regenerate the neighborhood database on a point-by-point basis is infeasible, and so again some temporal extent for building that database must be defined.

In the work presented here, the projection of nonstationary signals such as speech waveforms is done using 50% overlap Hanning windows. For global projection, the entire window is projected and an enhanced signal is reconstructed with weighted time-alignment, then the enhanced windows are recombined using the overlap-add technique. For local projection, we have created an approach that we call the “double-windowing method”, where an outer window is used to build a neighborhood region database, and then the signal within a centered inner window is projected point-by-point using its neighbors with respect to this database. This allows each point to be individually transformed and to have temporally appropriate neighbors, but without the large time complexity associated with searching for neighbors at each projection. The window is stepped to give a 50% overlap of the inner windows, which are recombined using the overlap-add technique. Block diagrams of the overall enhancement process for global and local projection methods are given in Fig. 1.

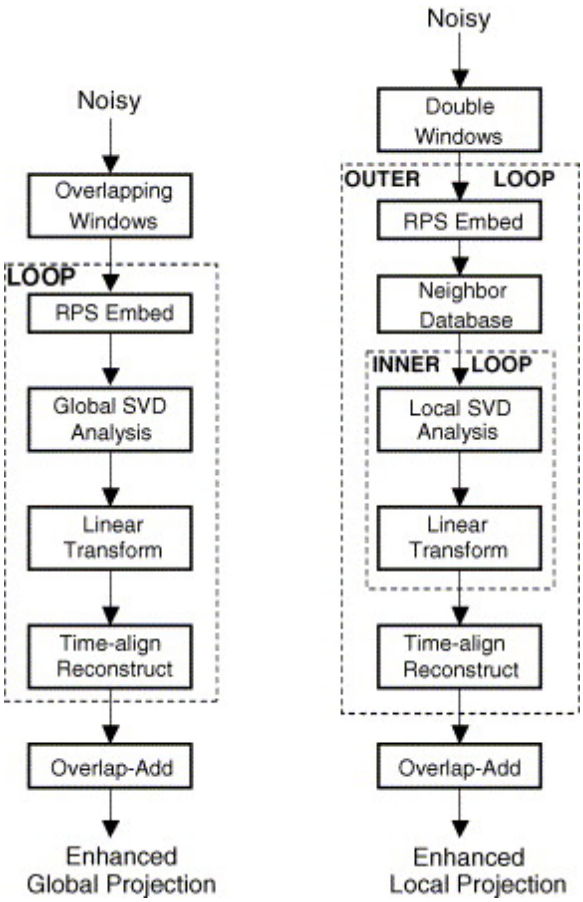

Fig. 1. Block diagrams of the global and local projection methods.

7. Experimental results

To illustrate the impact of altering the weighting matrix on projection methods, we have compared the performance of the $\mathbf{H}_{TDC}$ across values of μ ranging from 0.0 (the $\mathbf{H}_{LS}$ projection, a.k.a. Broomhead–King SSA projection) to 1.0 (the $\mathbf{H}_{LMMSE}$ projection). Experiments include global and local projection methods applied to data generated from the Henon map, given by:

$$x_{n+1} = 1 - ax_n^2 + bx_{n-1},$$

as well as to speech waveform data taken from the TIMIT data set [29].

To evaluate the results and illustrate the impact of moving from the LS to the LMMSE projections, we extend Eq. (17) for the purpose of calculating signal distortion and noise residual on RPS transformations. Using Eq. (25), the time-aligned weighted error signal can be decomposed as:

$$(26)\ \mathbf{e} = \hat{\mathbf{x}} - \mathbf{x} = \sum_i \mathbf{p}_i \cdot (\mathbf{H}_i - \mathbf{I})\mathbf{x} + \sum_i \mathbf{p}_i \cdot \mathbf{H}_i\mathbf{w} \triangleq \mathbf{e}_x + \mathbf{e}_w,$$

where $\mathbf{H}_i$ is the ith row of the transform matrix and $\mathbf{x}$ and $\mathbf{w}$ the original signal and added noise signal, respectively. The essence of (26) is that total signal distortion and total noise residual can be determined using the same weighted time-aligned sum as the signal, with modified projection operators $(\mathbf{H} - \mathbf{I})\mathbf{x}$ and $\mathbf{Hw}$, respectively. We can then use mean-square signal distortion $\|\mathbf{e}_x\|$, mean-square noise residual $\|\mathbf{ew}\|$, and mean-square total error $\|\mathbf{e}\|$ as metrics to evaluate the noise reduction process for a particular set of implementation parameters. To eliminate scaling effects, the results given here are normalized by dividing these quantities by

the energy of the clean signal $\|\mathbf{x}\|$. In addition, along with normalized total error we include the common metric of signal-to-noise ratio (SNR):

(27) $\text{SNR} = 10\log_{10}\left(\frac{\sum x_n^2}{\sum (x_n - \hat{x}_n)^2}\right) = 10\log_{10}\left(\frac{\|\mathrm{x}\|}{\|\mathrm{e}\|}\right)$.

The original SNR for all examples is 0 dB, so that the final SNR also indicates net improvement.

Parameters used for the Henon map are a = 1.4, b = 0.3 (with corresponding Lyapunov exponents of $\lambda \approx 0.42$, −1.62). For the Henon map, there is no windowing, i.e. a single window is used for the entire 3000-point signal (an additional 500 points of the signal are discarded to ensure signal convergence to the attractor). The initial embedding dimension is 5 for global projection and 20 for local projection, with a final projection dimension of 3 for both. Ten signals are generated with different initial conditions, and the results are averaged across the examples.

For the speech data, the initial embedding dimension is 20, and the projection dimension is 5. Speech windows are 256 points, with 128 points overlap. Ten different examples are used, and the results averaged across examples as with the Henon map. For both the Henon map and the speech data, the neighborhood region of a point is defined as the set of the nearest 25 points in the trajectory matrix.

Results are shown in Table 1, Table 2, using the algorithms outlined in Fig. 1 and the $\mathbf{H}_{\text{TDC}}$ projection from Eq. (22), with additive white noise at 0 dB SNR. The evaluation metrics include normalized mean square signal distortion $\|\mathbf{e}_x\|/\|\mathbf{x}\|$, normalized mean square noise residual $\| \mathrm{e}_w \|/\| \mathrm{x} \|$, normalized mean square total error $\|\mathbf{e}\|/\|\mathbf{x}\|$, and SNR, as outlined above.

Table 1. Projection results for Henon map with 0 dB SNR additive white Gaussian noise

μ	**Global projection d = 5 → 3**				**Local projection d = 20 → 3**			
	Normalized signal distortion $\|e_x\|/\|x\|$	**Normalized noise residual** $\|ew\|/\|x\|$	**Normalized total error $\|e\|/\|x\|$**	**SNR (dB)**	**Normalized signal distortion $\|e_x\|/\|x\|$**	**Normalized noise residual** $\|ew\|/\|x\|$	**Normalized total error $\|e\|/\|x\|$**	**SNR (dB)**
0.0	.188	.409	.598	1.24	.395	.131	.434	3.63
0.1	.262	.238	.498	3.03	.378	.127	.414	3.84
0.2	.343	.161	.505	2.98	.393	.128	.427	3.70
0.3	.382	.131	.518	2.86	.395	.126	.430	3.66
0.4	.436	.101	.533	2.74	.387	.134	.423	3.70
0.5	.486	.077	.561	2.51	.385	.130	.421	3.76
0.6	.503	.070	.574	2.42	.398	.131	.432	3.65
0.7	.523	.063	.586	2.32	.394	.139	.428	2.69
0.8	.553	.053	.603	2.20	.390	.130	.423	3.74
0.9	.586	.045	.632	2.00	.392	.126	.434	3.73
1.0	.585	.044	.628	2.02	.398	.129	.429	3.68

Table 2. Projection results for speech signals with 0 dB SNR additive white Gaussian noise

μ	**Global projection d = 20 → 3**				**Local projection d = 20 → 3**			
	Normalized signal distortion $\|e_x\|/\|x\|$	**Normalized noise residual** $\|ew\|/\|x\|$	**Normalized total error $\|e\|/\|x\|$**	**SNR (dB)**	**Normalized signal distortion $\|e_x\|/\|x\|$**	**Normalized noise residual** $\|ew\|/\|x\|$	**Normalized total error $\|e\|/\|x\|$**	**SNR (dB)**
0.0	.037	.239	.255	6.37	.150	.238	.373	4.40
0.1	.039	.191	.214	6.95	.149	.232	.367	4.46
0.2	.041	.163	.190	7.34	.149	.232	.366	4.48
0.3	.045	.142	.176	7.62	.150	.230	.364	4.49
0.4	.046	.124	.161	7.92	.150	.227	.361	4.52
0.5	.050	.111	.153	8.10	.153	.224	.361	4.54
0.6	.052	.103	.148	8.23	.152	.223	.358	4.54
0.7	.055	.095	.143	8.34	.152	.220	.355	4.57

0.8	.056	.090	.140	8.45	.154	.221	.357	4.55
0.9	.060	.082	.136	8.58	.155	.216	.353	4.58
1.0	.063	.078	.135	8.62	.156	.218	.356	4.54

8. Discussion

In all cases, the change in balance between signal distortion and noise residual components of the overall error is evident as μ moves from a LS projection at 0.0 to an LMMSE projection at 1.0. As would be expected this change is much more substantial for the global projection method than for local projection, since the smaller neighborhood regions are reasonably well approximated by a linear model. The overall error is given by a combination of the signal distortion and noise residual, and the best value for μ is a function of the minima, maxima, and rate of change for each of these components as μ is varied. With the speech data, the noise residual component is dominant, and overall error decreases as μ increases. With the Henon map, the two components have different rates of change with respect to μ, and the overall error peaks at about $\mu = 0.1$, at which point the noise residual has already dropped significantly but signal distortion has not yet increased as much. In general, the adjustment of this parameter to balance distortion and residual may allow for an overall improvement with respect to total error for noisy systems.

There are a number of interesting contrasts between the Henon system task and the speech enhancement task. The most clearly evident of these is that for the Henon task, where the signal in question is generated by a chaotic dynamical system, the local projection approach to enhancement gives better results, whereas for the speech task, where the signals are generated by a mechanism that is relatively well-approximated by linear systems, the global projection approach offers substantially better noise reduction. Another contrast is that the overall degree of change in distortion/residual balance as μ changes is measurably greater for the speech task.

9. Conclusions

We have presented a new approach for generalizing the concept of phase space projection, which can be viewed on a global level as an extension of Broomhead–King phase space reconstruction. The approach, based on a subspace decomposition approach taken from the statistical signal processing field, has the advantages that it incorporates knowledge of the underlying noise statistics into the projection mechanism and that it grants explicit control over the balance between the signal distortion and noise residual components of the remaining error. Example results on both known dynamical systems and collected signal data corrupted with additive white noise illustrate and support the use of this generalized approach.

References

[1] E. Ott. **Chaos in Dynamical Systems.** Cambridge University Press, Cambridge, UK (1993)

[2] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw. **Geometry from a time series.** Phys. Rev. Lett., 45 (1980), pp. 712-716

[3] H. Whitney. **Differentiable manifolds.** The Annals of Mathematics, 2nd Series, vol. 37 (1936) pp. 645–680

[4] F. Takens. **Detecting strange attractors in turbulence.** Proceedings of the Dynamical Systems and Turbulence, Warwick (1980), pp. 366-381

[5] T. Sauer, J.A. Yorke, M. Casdagli. **Embedology.** J. Stat. Phys., 65 (1991), pp. 579-616

[6] E. Bradley. **Analysis of time series.** M. Berthold, D. Hand (Eds.), An Introduction to Intelligent Data Analysis, Springer, New York (1999), pp. 167-194

[7] H. Kantz, T. Schreiber. **Nonlinear Time Series Analysis.** Cambridge University Press, Cambridge (1997)

[8] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring. **The analysis of observed chaotic data in physical systems.** Rev. Mod. Phys., 65 (1993), pp. 1331-1392

[9] D. Ruelle. **Elements of Differentiable Dynamics and Bifurcation Theory.** Academic Press, New York (1989)

[10] G.L. Baker, J.P. Gollub. **Chaotic Dynamics.** Cambridge University Press, New York (1990)

[11] H.D.I. Abarbanel. **Analysis of Observed Chaotic Data.** Springer, New York (1996)

[12] J.F. Gibson, J.D. Farmer, M. Casdagli, S. Eubank. **An analytic approach to practical state space reconstruction.** Physica D, 57 (1992), pp. 1-30

[13] D.S. Broomhead, G. King. **Extracting qualitative dynamics from experimental data.** Physica D (1986), pp. 217-236
[14] T. Schreiber, P. Grassberger. **A simple noise-reduction method for real data.** Phys. Lett. A, 160 (1991), pp. 411-418
[15] R. Cawley, G.-H. Hsu. **SNR performance of a noise reduction algorithm applied to coarsely sampled chaotic data.** Phys. Lett. A, 166 (1992), pp. 188-196
[16] T. Sauer. **A noise reduction method for signals from nonlinear systems.** Physica D, 58 (1992), pp. 193-201
[17] S.M. Hammel. **A noise reduction method for chaotic systems.** Phys. Lett. A, 148 (1990), pp. 421-428
[18] J.D. Farmer, J.J. Sidorowich. **Optimal shadowing and noise reduction.** Physica D, 47 (1991), pp. 373-392
[19] R. Bowen. **ω-limit sets for axiom A diffeomorphisms.** J. Differ. Equ., 18 (1975), pp. 333-339
[20] L.L. Sharf. **Statistical Signal Processing: Detection, Estimation, and Time-Series Analysis.** Addison-Wesley, New York (1990)
[21] D.G. Manolakis, V.K. Ingle, S.M. Kogon. **Statistical and Adaptive Signal Processing.** McGraw-Hill (2000)
[22] Y. Ephraim, H.L.V. Trees. **A signal subspace approach for speech enhancement.** IEEE Trans. Speech Audio Process., 3 (1995), pp. 251-266
[23] H.L. Van Trees. Detection, Estimation, and Modulation Theory, vol. I, Wiley, New York (1968)
[24] L. Scharf. **Statistical Signal Processing.** Prentice-Hall, Englewood Cliffs, NJ (1991)
[25] J. Rissanen. **Modeling by shortest data description.** Automatica, 14 (1978), pp. 465-471
[26] M. Wax, T. Kailath. **Detection of signals by information theoretic criteria.** IEEE Trans. Acoust. Speech Signal Process., 33 (1985), pp. 387-392
[27] N. Merhav. **The estimation of the model order in exponential families.** IEEE Trans. Inform. Theory, 35 (1985), pp. 1109-1114
[28] C. Merkwirth, U. Parlitz, W. Lauterborn. **Fast nearest-neighbor searching for nonlinear signal processing.** Phys. Rev. E, 62 (2000), p. 2089
[29] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, V. Zue, TIMIT Acoustic-Phonetic Continuous Speech Corpus: Linguistic Data Consortium, 1993.