4

Additional Theory and Applications of the KL Expansion

In this Chapter we continue our study of the K-L procedure and apply it in detail to a variety of problems, including the study of patterns which evolve in time, i.e., spatio-temporal data.

In Section 4.1 we develop the continuous version of the KL transform. This is important for theoretical reasons and will be used later Sections. Section 4.2 develops the KL procedure for both continuous and discrete vector functions. In Section 4.3 we demonstrate how both the continuous and discrete expansions may be extended to have even and odd symmetry. The appliction of the methodology to the reduction of dynamical equations is presented in Section 4.4. In Section 4.5 we present an extension of the KL procedure for *gappy* data. This is followed the application of the KL procedure in the presence of noise in Section 4.6. The last major Section of this comprehensive Chapter on optimal expansions develops the local KL procedure. Of particular importance in Section 4.7 is the application of the method to computing local dimensionalities using the scaling properties of the singular values.

4.1 THE CONTINUOUS KL TRANSFORM

In many cases of importance it is easier to demonstrate theoretical results concerning the Karhunen-Loève procedure if the continuous form of the expansion is used. The derivation now moves to a function space setting, but remains essentially analogous to the derivation for the discrete case.

Now we assume that the data points $\{u^{(\mu)}(x)\}_{\mu=1}^{P}$ of our ensemble are functions which reside in the Hilbert space $L^2(a,b)$, $\infty \leq a < b \leq \infty$, i.e., the space of Lebesgue measurable functions

$$f : (a,b) \to \mathbb{C}$$

which are square-integrable

$$\int_a^b |f(x)|^2 dx < \infty$$

with the inner product

$$(f,g) = \int_a^b f(x)\overline{g(x)}dx \tag{4.1}$$

and induced norm

$$\|f\|^2 = (f,f).$$

The presentation here will not be technically exacting in the sense that sets of measure zero will be ignored. For a more mathematically detailed approach see [74]. In addition, we will assume that all of our functions are real.

In this infinite dimensional inner product space, we again seek to construct an optimal basis $\mathcal{B}$ with elements $\{\phi^{(j)}(x)\}_{j=1}^{\infty}$ in L^2. Our basis should be

o.n., i.e.,

$$(\phi^{(i)}, \phi^{(j)}) = \int_a^a \phi^{(i)}(x)\phi^{(j)}(x)dx = \delta_{ij}.$$

Then, any square-integrable function can be expanded as

$$u^{(\mu)}(x) = \sum_{i=1}^{\infty} a_i^{(\mu)} \phi^{(i)}(x)$$

where the expansion coefficients $a_i^{(\mu)}$ are given by

$$a_i^{(\mu)} = (u^{(\mu)}, \phi^{(i)}) = \int_{-\infty}^{\infty} u^{(\mu)}(x)\phi^{(i)}(x)dx.$$

The ensemble average is defined as before, namely

$$\langle u(x) \rangle = \frac{1}{P} \sum_{\mu=1}^{P} u^{(\mu)}(x)$$

If the functions are time dependent, i.e., they are of the form $u(x,t)$ there is a related time-average defined as

$$\langle u(x) \rangle = \frac{1}{T} \int_0^T u(x,t)dt.$$

If the time-dependent function is sampled discretely in time we may collect an ensemble as before where

$$u^{(\mu)}(x) = u(x, t_\mu).$$

Often the approximation is made that the ensemble average and the time-average are equal. If this is true the flow is said to be *ergodic* [70].

As before, we proceed by defining the first basis function $\phi^{(1)}(x)$ by means of an optimization criterion. The mean-square projection of the data onto this function should be a maximum:

$$\langle (\int_{-\infty}^{\infty} u(x)\phi^{(1)}(x)dx)^2 \rangle = \text{maximum}$$

subject to

$$\int_{-\infty}^{\infty} (\phi^{(1)}(x))^2 dx = 1.$$

The remaining basis functions may be defined proceeding sequentially

$$\max_{\phi^{(j)}} \langle (\phi^{(j)}, u)^2 \rangle$$

subject to the side constraints

$$(\phi^{(j)}, \phi^{(k)}) = \delta_{jk}, \text{ for } k < j$$

Using the technique of Lagrange multipliers and the calculus of variations it can then be shown that the basis functions are solutions to the integral equation

$$\int C(x,y)\phi(y)dy = \lambda\phi(x) \tag{4.2}$$

where

$$C(x,y) = \langle u(x)u(y)\rangle \tag{4.3}$$

The type of this integral equation is known as a Fredholm Equation of the 2nd Kind and its properties fall within the scope of Hilbert-Schmidt theory [29]. This integral equation is seen to be the continuous analogue to the eigenvector problem for obtaining a best basis in a finite-dimensional setting. Among the properties of the solutions of this *eigenfunction* problem are the following:

- The kernel is symmetric, i.e., $C(x,y) = C(y,x)$.
- The eigenvalues are real, countable, and non-negative.
- The total energy is finite, i.e., $\sum_n \lambda_n < \infty$.
- The eigenfunctions form an orthonormal basis for the Hilbert space L^2.

In addition, we have the continuous analogue to the spectral theorem

Theorem 4.1. *Given $C(x,y) = \langle u(x)u(y)\rangle$ is continuous in x, y then*

$$C(x,y) = \sum_i \lambda_i \phi^{(i)}(x)\phi^{(i)}(y)$$

where the series converges uniformly and absolutely to $C(x,y)$.

This is known as Mercer's theorem. We refer the reader to [74, 66] for further details.

Returning to the data analysis problem, it is possible that the relative frequency $p^{(\alpha)}$ of each member of the ensemble is known, or may be estimated. If the ensemble consists of P elements, then $Pp^{(\alpha)}$ of the elements are indexed by α. The relative frequency satisfies the conditions

$$p^{(\alpha)} \geq 0, \quad \sum_{\alpha=1}^{P} p^{(\alpha)} = 1$$

The *weighted* covariance matrix is then defined as

$$C(x,y) = \sum_{\alpha=1}^{P} p^{(\alpha)} u(x) u(y) \tag{4.4}$$

This weighted covariance matrix may then used in place of the evenly weighted ensemble average covariance matrix.

In the next Section we see how the continuous transform allows us to extend these ideas to vector functions. Furthermore, we show that discrete data made up of concatenated vectors derived from continuous functions can be dealt with in a computationally efficient manner.

4.2 VECTOR FUNCTION KL EXPANSIONS

One of the most important applications of the KL expansion is to data consisting of several variables with values defined over a mulit-dimensional domain. This section deals with the extension of the continuous KL expansion to such vector functions. The procedure is also extended to the fully discrete setting.

4.2.1 Continuous Vector Functions

Here we address the application of the KL procedure to an ensemble of vector functions of the form

$$\mathbf{u}^{(\mu)}(\mathbf{x}) = (u_1^{(\mu)}(\mathbf{x}), \ldots, u_K^{(\mu)}(\mathbf{x}))^T$$

where each

$$u_i : \mathbb{R}^j \to \mathbb{R}$$

and j is the dimension of the domain. Such "data sets" are difficult to generate in practice, but they arise naturally in theoretical settings.

Example 4.1. Consider the fluid flow with scalar flow variables consisting of the concatenated vector function

$$\mathbf{u}(\mathbf{x},t) = (u(\mathbf{x},t), v(\mathbf{x},t), w(\mathbf{x},t), e(\mathbf{x},t), \rho(\mathbf{x},t))^T$$

where $\mathbf{x} = (x_1, x_2, x_3)$ and u, v, w are the flow velocities in the x_1, x_2 and x_3 directions, respectively; e is the internal energy and ρ is the fluid density.

To determine a best basis for such vector functions we must extend the definition of the kernel of the integral equation. The appropriate extension is

$$\mathbf{C}(\mathbf{x},\mathbf{x}') = \langle \mathbf{u}(\mathbf{x})\mathbf{u}(\mathbf{x}') \rangle.$$

where

$$\mathbf{C}_{ij}(\mathbf{x},\mathbf{x}') = \langle u_i(\mathbf{x}) u_j(\mathbf{x}') \rangle.$$

The kernel $\mathbf{C}$ is now referred to as a two-point correlation tensor. If $K = 2$ we see that

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} = \begin{pmatrix} \langle u_1(\mathbf{x})u_1(\mathbf{x}')\rangle & \langle u_1(\mathbf{x})u_2(\mathbf{x}')\rangle \\ \langle u_2(\mathbf{x})u_1(\mathbf{x}')\rangle & \langle u_2(\mathbf{x})u_2(\mathbf{x}')\rangle \end{pmatrix}$$

The integral equation which produces the optimal eigenfunctions is now given by

$$\int \mathbf{C}(\mathbf{x}, \mathbf{x}')\boldsymbol{\phi}(\mathbf{x}')\mathbf{d}\mathbf{x}' = \lambda\boldsymbol{\phi}(\mathbf{x}).$$

This may be thought of $\boldsymbol{\phi}(\mathbf{x})$ as being a concatenation of components of eigenfunctions

$$\boldsymbol{\phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), \phi_2(\mathbf{x})).$$

The integral equation can also be written in component form as

$$\int \sum_{j=1}^{2} \mathbf{C}_{ij}(\mathbf{x}, \mathbf{x}')\phi_j(\mathbf{x}')\mathbf{d}\mathbf{x}' = \lambda\phi_i(\mathbf{x}).$$

for $i = 1, 2$.

The Snapshot Method

The snapshot method again helps us reduce the problem for degenerate kernels. We again start with the data-dependent representation

$$\boldsymbol{\phi}(\mathbf{x}) = \sum_{\mu=1}^{P} a_\mu \mathbf{u}^{(\mu)}(\mathbf{x})$$

and component-wise

$$\phi_j(\mathbf{x}) = \sum_{\mu=1}^{P} a_\mu u_j^{(\mu)}(\mathbf{x}).$$

Substituting this into the component integral equations leads to

$$\int \sum_j \langle u_i(\mathbf{x})u_j(\mathbf{x}')\rangle (\sum_\mu a_\mu u_j^{(\mu)}(\mathbf{x}'))\mathbf{d}\mathbf{x}' = \lambda \sum_\nu a_\nu u_i^{(\nu)}(\mathbf{x}).$$

Expanding the ensemble average

$$\int \sum_j (\frac{1}{P}\sum_\nu u_i^{(\nu)}(\mathbf{x})u_j^{(\nu)}(\mathbf{x}'))(\sum_\mu a_\mu u_j^{(\mu)}(\mathbf{x}'))\mathbf{d}\mathbf{x}' = \lambda \sum_\nu a_\nu u_i^{(\nu)}(\mathbf{x}).$$

Rearranging,

$$\sum_\nu u_i^{(\nu)}(\mathbf{x})[\sum_\mu a_\mu \sum_j \int u_j^{(\nu)}(\mathbf{x}')u_j^{(\mu)}(\mathbf{x}')\mathbf{d}\mathbf{x}' - (\lambda P)a_\nu] = 0.$$

In other words,

$$\sum_{\nu} u_i^{(\nu)}(\mathbf{x})[\sum_{\mu} \mathbf{L}_{\nu\mu} a_\mu - \lambda P a_\nu] = 0$$

where

$$\mathbf{L}_{\nu\mu} = \int \sum_j u_j^{(\nu)}(\mathbf{x}') u_j^{(\mu)}(\mathbf{x}') \mathbf{dx}'.$$

Hence,

$$\mathbf{La} = \tilde{\lambda}\mathbf{a}$$

where we have used $\tilde{\lambda} = \lambda P$.

4.2.2 Discrete Vector Functions

As we have already observed, pattern data typically comes not in the form of continuous functions but rather as discrete vectors. The discussion in Section 3.4 dealt directly with the application of the KL procedure to the case of a single independent discrete variable. The results of the previous Section apply to continuous vector functions. We now modify these results so that we can treat discretely sampled vector data.

Example 4.2. A typical source of discrete multivariable vector data is a numerical simulation of physical problem, such as the motion of a fluid. The continuous data described in Example 4.1 is now estimated on a 3-dimensional lattice, or grid. For example, the first component of velocity $u(\mathbf{x}, t)$ is computed on $u(x_1, x_2, x_3, t)$ where $x_i \in \{1, \ldots, N_i\}$. This data may be concatenated into a single column vector as

$$\mathbf{u}(t) = \begin{pmatrix} u(1,1,1) \\ \vdots \\ u(N_1, N_2, N_3) \end{pmatrix}$$

Example 4.3. A color digital image is commonly given as a triplet of red, $r(i,j)$, green, $g(i,j)$ and blue $b(i,j)$ color values on a two-dimensional lattice. Concatenating the rows (or columns) of each image, and then the images leads to the high-dimensional vector

$$\mathbf{x} = (\mathbf{r}, \mathbf{g}, \mathbf{b})^T.$$

Lets consider the application of the above where the continuous vector function is actually discretized so that

$$\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)^T$$

where

$$\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n \quad \mathbf{u} \in \mathbb{R}^{2n}.$$

Then

$$\begin{aligned}\mathbf{C} &= \langle \mathbf{u}\mathbf{u}^T \rangle \\ &= \begin{pmatrix} \langle \mathbf{u}_1\mathbf{u}_1^T \rangle & \langle \mathbf{u}_1\mathbf{u}_2^T \rangle \\ \langle \mathbf{u}_2\mathbf{u}_1^T \rangle & \langle \mathbf{u}_2\mathbf{u}_2^T \rangle \end{pmatrix}\end{aligned}$$

Thus each $\mathbf{C}_{ij}$ is an $n \times n$ matrix and $\mathbf{C}$ is a $2n \times 2n$ matrix. Again we have the eigenvector problem

$$\mathbf{C}\phi = \lambda\phi$$

which leads to the snapshot equation, as in the continuous case considered in the previous Section, i.e.,

$$\mathbf{L}\mathbf{a} = \lambda P\mathbf{a}$$

where $\mathbf{L}$ is a $P \times P$ matrix and $\mathbf{L}_{\nu\mu} = \sum_j (\mathbf{u}_j^{(\nu)}, \mathbf{u}_j^{(\mu)})$, i.e., a vector dot product.

4.3 SYMMETRIC OPTIMAL EIGENFUNCTIONS

Historically, one of the most important eigenfunction expansion is the continuous Fourier series

$$f(x) = \sum_k a_k \cos(kx) + b_k \sin(kx) \tag{4.5}$$

which decomposes a periodic function into its even and odd parts. The $\{a_k\}$ coefficients represent the odd portion of $f(x)$ while the $\{b_k\}$ represent the even. It is easy to show using the orthogonality of the sinusiods that if $f(x)$ is even, then the $\{b_k\}$ are all zero, and if $f(x)$ is odd, then the $\{a_k\}$ are all zero. This property often simplifies the computations associated with using the Fourier series expansion.

Recall that a function $f(x)$ is said to be even if $f(x) = f(-x)$ and odd if $f(x) = -f(-x)$. In addition, any function may be expressed as the sum of an even and odd function using the identity

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}.$$

For example,

$$e^x = \frac{e^x + e^{-x}}{2} + \frac{e^x - e^{-x}}{2}$$

See Figure 4.1 for the plots of the individual functions in the decomposition.

The decomposition of a function into even and odd orthogonal components may be generalized to the form

$$f(x) = \sum_k a_k f_e^{(k)}(x) + b_k f_o^{(k)}(x) \tag{4.6}$$

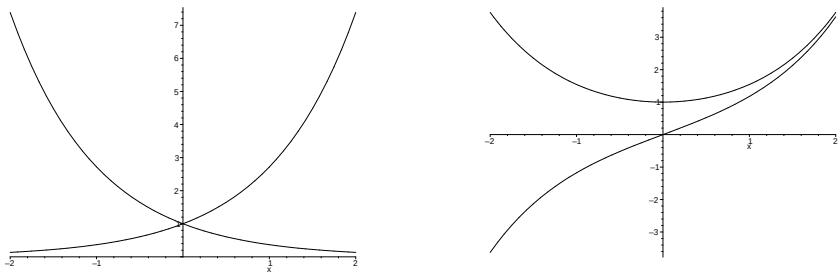

Fig. 4.1 Left: the function e^x and its reflection e^{-x}. Right: The functions on the left may be decomposed in terms of the even and odd functions $f_e(x) = (e^x + e^{-x})/2, f_o(x) = (e^x - e^{-x})/2$.

where $\{f_e^{(k)}(x)\}$ are even functions, $\{f_o^{(k)}(x)\}$ are odd functions, and together they form a basis for the function space in question. Given such a decomposition it is possible to characterize the symmetry of a pattern or an ensemble of patterns. For instance, the degree of *evenness* of an ensemble may be quantified by the sum $\sum_k \langle a_k^2 \rangle$ and degree of *oddness* as $\sum_k \langle b_k^2 \rangle$. The symmetric KL procedure will automatically produce these quantities as eigenvalues.

Given the convenience of such a decomposition for characterizing the symmetric components of a function (or data vector) it is natural to address the issue of symmetric optimal eigenfunction (or eigenvector) expansions.

In general, if $C(x, y) = \langle u(x)u(y) \rangle$ and the optimal eigenfunctions in the expansion are determined by solving the equation

$$\int C(x, y)\phi(y)dy = \lambda\phi(x) \tag{4.7}$$

then the eigenfunctions $\{\phi^{(j)}\}$ are neither even nor odd, i.e., the possess no symmetry. As a consequence, the optimal eigenfunction expansion

$$u(x) = \sum_{j=1}^{\infty} a_j \phi^{(j)}(x)$$

does not permit the splitting of the decomposition into even and odd subspaces.

The topic of this Section is to demonstrate how a simple modification to the KL procedure permits the construction of optimal bases with even and odd

eigenfunctions. Although the initial setting for the discussion is the continuous transform, all of the ideas carry over to the discrete case.

We begin by defining the reflection operator

$$R\phi(x) = \phi(-x).$$

Now consider an ensemble of patterns $\{u^{(\mu)}(x)\}$ with $\mu = 1, \ldots, P$.

Definition 4.1. *The symmetry extended ensemble is defined as the union*

$$\tilde{X} = \{u^{(\mu)}(x)\} \cup \{Ru^{(\mu)}(x)\}$$

where $\mu = 1, \ldots, P$.

The extended ensemble average is then given by

$$\langle u(x) \rangle = \frac{1}{2P} \sum_{\mu=1}^{P} (u^{(\mu)}(x) + Ru^{(\mu)}(x)).$$

It is interesting to note that the extended ensemble average is an even function. As usual, we form the fluctuating ensemble

$$\tilde{u}^{(\mu)}(x) = u^{(\mu)}(x) - \langle u(x) \rangle$$

and for simplicity we again immediately drop the tilde notation.

It will be demonstrated that the symmetric optimal eigenfunction expansion is found by solving the *symmetrized* integral equation

$$\int \hat{C}(x,y)\hat{\phi}(y)dy = \lambda\hat{\phi}(x) \tag{4.8}$$

where the symmetrized kernel $\hat{C}$ is given by

$$\hat{C}(x,y) = \frac{1}{2P} \sum_{\mu=1}^{P} (u^{(\mu)}(x)u^{(\mu)}(y) + Ru^{(\mu)}(x)Ru^{(\mu)}(y)). \tag{4.9}$$

Now we propose to compare the solutions of the symmetrized integral Equation (4.8) with those of the original eigenfunction problem of Equation (4.7). In particular, it will be shown that the eigenfunctions of Equation (4.8) are symmetric, i.e., even and odd functions. In addition, it will be shown that the solutions of the symmetrized integral equation may be found by solving associated even and odd integral equations.

The set of eigenfunctions which satisfy a given integral equation also define an eigenspace, in a manner analogous with the discrete eigenvector problem. For example, the eigenspace associated with the symmetrized kernel $\hat{C}$ is the set of functions

$$E(\hat{C}) = \{\hat{\phi} : \int \hat{C}(x,y)\hat{\phi}(y)dy = \lambda\hat{\phi}(x)\} \tag{4.10}$$

The analysis of Equation (4.8) will be facilitated by the introduction of the even functions

$$u_e^{(\mu)}(x) = \frac{u^{(\mu)}(x) + Ru^{(\mu)}(x)}{2}$$

and the associated odd functions

$$u_o^{(\mu)}(x) = \frac{u^{(\mu)}(x) - Ru^{(\mu)}(x)}{2}.$$

For $\mu = 1, \ldots, P$ we may view these functions as defining an even ensemble $\{u_e^{(\mu)}(x)\}$ and an odd ensemble $\{u_o^{(\mu)}(x)\}$.

These ensembles also define kernels. The even kernel

$$C_e(x, y) = \langle u_e(x) u_e(y) \rangle$$

is even in x and y and the odd kernel

$$C_o(x, y) = \langle u_o(x) u_o(y) \rangle$$

is odd in x and y.

Proposition 4.1. *The symmetrized kernel $\hat{C}$ may be decomposed into even and odd components using these kernels, i.e.,*

$$\hat{C}(x, y) = C_e(x, y) + C_o(x, y)$$

This splitting of the kernel into its even and odd components will permit us to establish that the integral equations which produce the even and odd eigenfunctions are the *even* integral equation

$$\int C_e(x, y) \phi_e(y) dy = \lambda \phi_e(x) \tag{4.11}$$

and the *odd* integral equation

$$\int C_o(x, y) \phi_o(y) dy = \lambda \phi_o(x) \tag{4.12}$$

respectively.

Theorem 4.2. *If $\hat{\phi}$ is an eigenfunction of the symmetrized kernel $\hat{C}$, i.e.,$\hat{\phi} \in E(\hat{C})$ and $\hat{\phi}(x) = \hat{\phi}_e(x) + \hat{\phi}_o(x)$ decomposes $\hat{\phi}$ into even and odd functions, then*

- $\hat{\phi}_e(x) \in E(C_e)$
- $\hat{\phi}_o(x) \in E(C_o)$

In follows that,

$$E(\hat{C}) \subseteq E(C_e) + E(C_o) \tag{4.13}$$

Proof. By assumption

$$\int \hat{C}(x,y)\hat{\phi}(y)dy = \lambda\hat{\phi}(x)$$

Decomposing into the even and odd components gives

$$\int (C_e(x,y) + C_o(x,y))(\hat{\phi}_e(y) + \hat{\phi}_o(y))dy = \lambda(\hat{\phi}_e(x) + \hat{\phi}_o(x))$$

After multiplying out and setting the appropriate terms to zero (see Exercise 4.7) we have

$$\int C_e(x,y)\hat{\phi}_e(y)dy + \int C_o(x,y)\hat{\phi}_o(y)dy = \lambda(\hat{\phi}_e(x) + \hat{\phi}_o(x))$$

Equating the even and odd parts produces Equations (4.11) and (4.12), respectively. □

The next theorem essentially says that the even and odd eigenfunctions are actually solutions to the symmetrized integral equation. In other words, $E(C_e) \subset E(\hat{C})$ and $E(C_o) \subset E(\hat{C})$.

Theorem 4.3. *If $\phi_e \in E(C_e)$, then $\phi_e \in E(\hat{C})$. Similarly, if $\phi_o \in E(C_o)$, then $\phi_o \in E(\hat{C})$.*

Proof. By assumption

$$\int C_e(x,y)\phi_e(y)dy = \lambda\phi_e(x)$$

So $\int (C_e(x,y) + C_o(x,y))\phi_e(y)dy = \lambda\phi_e(x)$ since the additional term is just zero. Hence it follows

$$\int \hat{C}(x,y)\phi_e(y)dy = \lambda\phi_e(x)$$

The proof for ϕ_o is analogous. □

From these results we may conclude that any solution ϕ_e of (4.11) or ϕ_o of (4.12) is a solution to (4.8) and that the even and odd eigfunctions span orthogonal subspaces. Furthermore, it also follows that

$$E(C_e) + E(C_o) \subseteq E(\hat{C})$$

Thus, by combining the statements of theorems 4.13 and 4.13, we conclude that

$$E(C_e) + E(C_o) = E(\hat{C})$$

It is also easy to show that the even and odd kernels produce orthogonal eigenfunctions.

Proposition 4.2. *The even and odd eigenfunctions are orthgonal, i.e.,*

$$E(C_e) \perp E(C_o)$$

The proof of this is left for the exercises.

In other words $E(\hat{C})$, the eigenspace determined by equation (4.8) can be expressed as the direct sum of $E(C_e)$ and $E(C_o)$; i.e.,

$$E(\hat{C}) = E(C_e) \dot{\oplus} E(C_o).$$

Thus to determine the even eigenfunctions $\{\hat{\phi}_e(y)\}$, and the odd eigenfunctions $\{\hat{\phi}_o(y)\}$ of $\hat{C}$, we may solve the "smaller" problems given by Equations (4.11) and (4.12).

This theorem shows that the eigenfunctions of the symmetrized integral Equation (4.8) are actually all either even or odd.

Theorem 4.4. *Assume that all eigenvalues of the symmetrized integral equation are distinct. If $\hat{\phi}$ is an eigenfunction of $\hat{C}$, then either $\hat{\phi} = \hat{\phi}_e$, or $\hat{\phi} = \hat{\phi}_o$.*

Proof. Since $\hat{\phi}$ is an eigenvector of $\hat{C}$ and $\hat{\phi} = \hat{\phi}_e + \hat{\phi}_o$ it follows

$$\int \hat{C}(x,y)(\hat{\phi}_e(y) + \hat{\phi}_o(y))dy = \lambda(\hat{\phi}_e(x) + \hat{\phi}_o(x)) \tag{4.14}$$

Also, by theorems 4.2 and 4.3, these eigenfunctions satisfy $\int \hat{C}(x,y)\hat{\phi}_e(y)dy = \lambda_e\hat{\phi}_e(x)$ and $\int \hat{C}(x,y)\hat{\phi}_o(y)dy = \lambda_o\hat{\phi}_o(x)$, or, after adding these equations together,

$$\int \hat{C}(x,y)(\hat{\phi}_e(y) + \hat{\phi}_o(y))dy = \lambda_e\hat{\phi}_e(x) + \lambda_o\hat{\phi}_o(x) \tag{4.15}$$

Equating the right-hand sides of Equations (4.15) and (4.14) leads to

$$\lambda(\hat{\phi}_e(x) + \hat{\phi}_o(x)) = \lambda_e\hat{\phi}_e(x) + \lambda_o\hat{\phi}_o(x)$$

or

$$(\lambda - \lambda_e)\hat{\phi}_e(x) + (\lambda_o - \lambda)\hat{\phi}_o(x) = 0$$

But $\hat{\phi}_e(x)$ and $\hat{\phi}_o(x)$ are independent so $\lambda = \lambda_e = \lambda_o$. But this is a contradiction, since the eigenvalues are distinct by assumption. The only remaining possibilities are

- $\hat{\phi}_e(x) = 0$, $\lambda = \lambda_o$, $\hat{\phi}(x) = \hat{\phi}_o(x)$
- $\hat{\phi}_o(x) = 0$, $\lambda = \lambda_e$, $\hat{\phi}(x) = \hat{\phi}_e(x)$

□

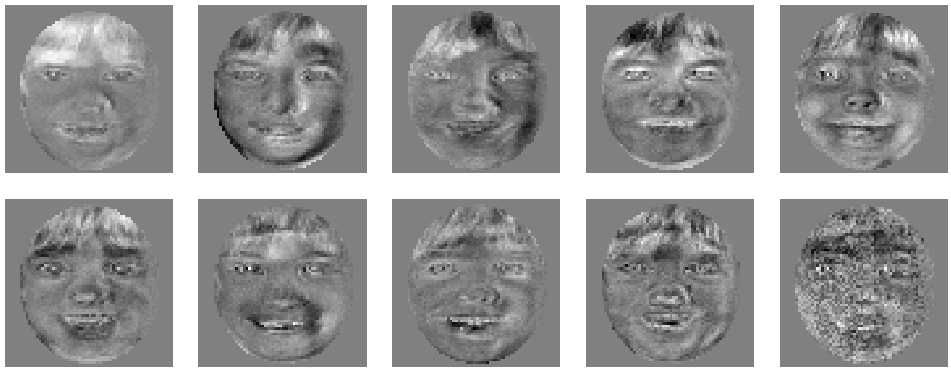

Fig. 4.2 The eigenvectors of a mean-subtracted ensemble of 10 faces.

4.3.1 Symmetric Optimal Eigenvectors

Although the continuous KL transform was a useful setting to derive the symmetric properties of the eigenfunctions for a symmetry extended data set, the discrete formulation is used in practical computations. This section outlines the discrete procedure which is analogous to the continuous KL of the previous section.

Definition 4.2. *A vector* $\mathbf{x} \in \mathbb{R}^N$ *is said to be even if*

$$x_i = x_{N-i+1}$$

and odd if

$$x_i = -x_{N-i+1}$$

for $i = 1, \ldots, N$.

Following the notation of the previous section we define the reflection of a vector about its midpoint as

$$(R\mathbf{x})_i = x_{N-i+1} \tag{4.16}$$

Example 4.4. A vector of length 4 may be said to be even if it has the form (a, b, b, a) and odd if it has the form $(a, b, -b, -a)$. All vectors of length 4 may be decomposed into the sum of an even and odd vector using

$$\mathbf{x} = \frac{1}{2}\begin{pmatrix} x_1 + x_4 \\ x_2 + x_3 \\ x_2 + x_3 \\ x_1 + x_4 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} x_1 - x_4 \\ x_2 - x_3 \\ -x_2 + x_3 \\ -x_1 + x_4 \end{pmatrix} \tag{4.17}$$

Following the previous section, the symmetric eigenvectors are computed from the symmetrized eigenvector problem

$$\hat{C}\hat{\phi} = \lambda\hat{\phi} \tag{4.18}$$

where

$$\hat{C} = \frac{\langle \mathbf{u}\mathbf{u}^T \rangle + \langle R\mathbf{u}R\mathbf{u}^T \rangle}{2}$$

There are twice as many patterns in this ensemble, so for rank deficient problems it is computationally less expensive to solve the associated even and odd eigenvector problems

$$C_e\phi_e = \lambda\phi_e$$

and

$$C_o\phi_o = \lambda\phi_o$$

where C_e and C_o are defined as before. Again, the even and odd patterns are found using

$$\mathbf{x}_e = \frac{\mathbf{x} + R\mathbf{x}}{2} \quad \mathbf{x}_o = \frac{\mathbf{x} - R\mathbf{x}}{2}$$

where the action of R is defined by Equation (4.16).

The Snapshot Method

When the matrices C_o and C_e are singular it is again useful to employ the Snapshot Method. Specifically, representing the even and odd eigenvectors in terms of the even and odd data we have

$$\phi_o = \sum_\mu b_\mu \mathbf{x}_o^{(\mu)}, \quad \phi_e = \sum_\mu a_\mu \mathbf{x}_e^{(\mu)}$$

yields the two reduced problems

$$L^e\mathbf{a}^{(\mu)} = \lambda\mathbf{a}^{(\nu)},$$

$$L^o\mathbf{b}^{(\mu)} = \lambda\mathbf{b}^{(\nu)}$$

where $L^e_{\nu\mu} = (\mathbf{x}_e^{(\nu)}, \mathbf{x}_e^{(\mu)})$ and $L^o_{\nu\mu} = (\mathbf{x}_o^{(\nu)}, \mathbf{x}_o^{(\mu)})$. Now there are two $P \times P$ eigenvector problems rather than a single $2P \times 2P$ problem which arises if we solve Equation (4.18) via the Snapshot method. Again, the resolution of the patterns is a factor only in the computation of the dot products and the required memory. As a result, very high-resoultion images may present a practical problem even if the eigenvector problems can be solved.

Example 4.5. *The Symmetric Rogues Gallery Problem.* The representation of digital images in Section 3.6.1 may now be extended to include symmetry. As an example, we revisit the Rogues Gallery problem and compute the even and odd eigenpictures. The result of computing the eigenpictures of the even

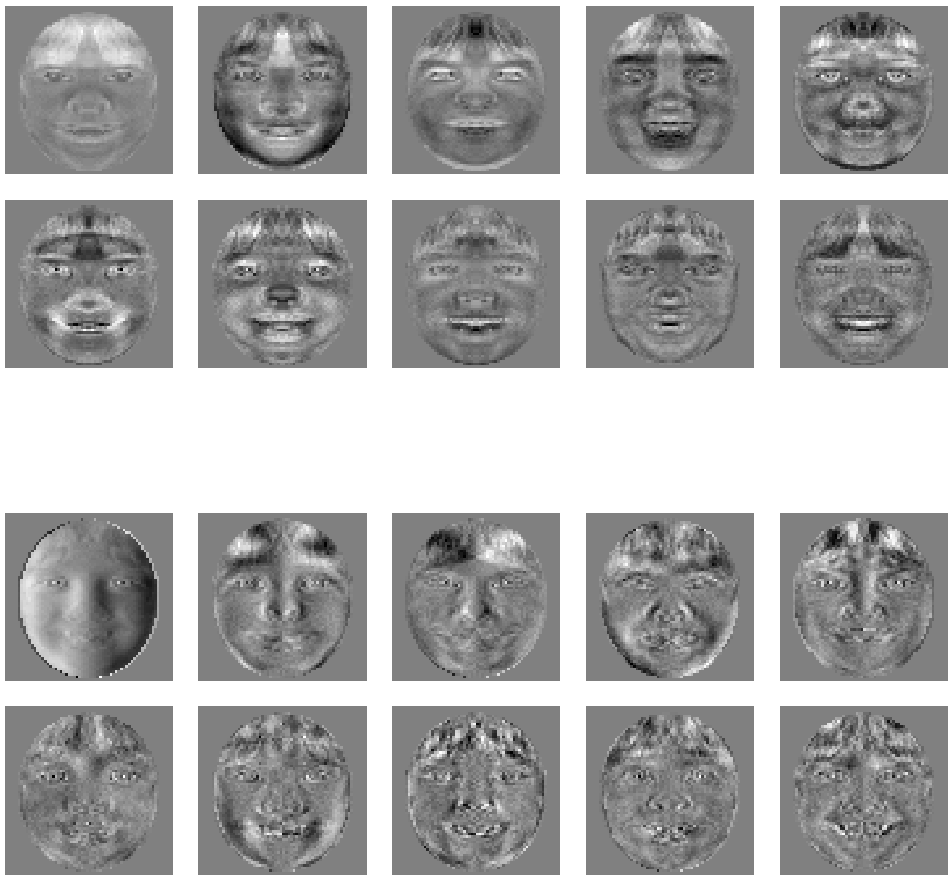

Fig. 4.3 Top: the even eigenvectors. Bottom: the odd eigenvectors.

ensemble is shown at the top of Figure 4.3; as expected, they are all even about the midline. The bottom of Figure 4.3 shows the result of computing the eigenpictures of the odd ensemble. Again, as expected, they are odd about the midline. For purposes of comparison the eigenpictures for the unextended ensemble are shown in Figure 4.2. See [45] for further details.

Although the presentation here considered only one type of symmetry, i.e., reflections, the ideas extend to other discrete and continuous symmetries. For several examples within the context of geometries of boundaries of fluid flows see [71]. For a mathematical treatment using representation theory see **??**.

4.4 LOW-DIMENSIONAL DYNAMICAL EQUATIONS

The modeling of complex phenomena such as atmospheric dynamics or ocean currents typically begins with systems of nonlinear partial differential equations. While the mathematical setting for such problems has an infinite number of dimensions, the actual computer implementation is always of finite dimension. In many instances this gap is not as wide as it first appears. For example, consider the Kuramoto-Sivashinsky equation (KS)

$$u_t + 4u_{xxxx} + \alpha(u_{xx} + \frac{1}{2}(u_x)^2) = 0 \tag{4.19}$$

where the subscripts denote partial differentiation, e.g., $u_x = \partial u/\partial x$. The KS equation is known, under certain conditions, to actually have solutions of finite dimension despite the infinite dimensional formulation [14]. In general, methods which lead to dimension estimates for solutions to PDEs are not constructive, i.e., they provide now means for developing model equations which reflect true dimension of the dynamics. The KL procedure, combined with standard techniques from numerical analysis, is a natural candidate for transforming infinite dimensional PDEs to systems of optimally low dimension. This idea was initially investigated in the 1950's in the context of numerical weather prediction [69]. In the interim this approach has received considerable attention; for a detailed theoretical treatment in the context of the Navier-Stokes equations see [70, 71, 72] and for a review of this area see [7].

Now we develop some of the basic ideas for building low-dimensional dynamical systems from partial differential equations. To begin, the standard Fourier-Galerkin is considered. This is followed by a discussion of a general projection procedure which employs the optimal KL basis. Lastly, a Sobolev, or derivative based norm especially suited for approximations of a dynamical nature is presented as an alternative to the usual mean-square error criterion.

4.4.1 The Galerkin Projection

The Galerkin Projection is a well-known method for generating a system of ordinary differential equations (ODEs) from a partial differential equation

(PDE) [14, 13]. For example, consider the PDE

$$\frac{\partial u}{\partial t} = \mathcal{N}(u)$$

where $\mathcal{N}$ is a nonlinear partial differential operator. Given a basis $\{\psi^{(n)}\}$ for the solutions to the PDE it follows

$$u(x,t) = \sum_{n=1}^{\infty} a_n(t)\psi^{(n)}(x). \tag{4.20}$$

The solution is then approximated by taking an D-term truncation of the solution

$$u_D(x,t) = \sum_{n=1}^{D} a_n(t)\psi^{(n)}(x). \tag{4.21}$$

Projecting the truncated solutions onto our basis as

$$(\psi^{(k)}, \frac{\partial u_D}{\partial t} - \mathcal{N}(u_D)) = 0 \tag{4.22}$$

produces a system of *amplitude* equations

$$\frac{da_k}{dt} = F_k(a_1, \ldots, a_D)$$

Before addressing the question of how to minimize D, we consider the Galerkin procedure using a sinusoidal basis.

4.4.2 The Fourier-Galerkin Projection

The Fourier modes $\{e^{inx}\}$ provide a useful basis for periodic phenomena. Periodic functions can be decomposed via the Fourier expansion

$$u(x,t) = \sum_{-\infty}^{\infty} a_n(t)e^{inx}. \tag{4.23}$$

The Fourier coefficients a_n are found via the orthogonality relationship

$$\int_0^{2\pi} e^{ikx}e^{-ijx} = 2\pi\delta_{jk}. \tag{4.24}$$

Substituting the Fourier expansion for $u(x,t)$ into the KS equation produces

$$\sum_{n=-\infty}^{\infty} (\dot{a}_n(t) + (4n^4 - \alpha n^2)a_n(t))e^{inx} - \frac{\alpha}{2}(\sum_{n=-\infty}^{\infty} na_n(t)e^{inx})^2 = 0 \tag{4.25}$$

Projecting this equation via Equation (4.22) onto the first D Fourier modes using the orthogonality condition results in

$$\dot{a}_l(t) = (\alpha l^2 - 4l^2)a_l(t) + \frac{\alpha}{2}\sum_{n=-N+l}^{N}(l-n)na_{l-n}a_n \tag{4.26}$$

where $-N \leq l \leq N$. Given $u(x,t)$ is a real-valued function these equations may be simplified by making use of the reality condition $a_l = \overline{a}_{-l}$. Note also that the a_0 term decouples from the system. The resulting coupled system is then

$$\dot{a}_l = l^2(\alpha - 4l^2)a_l + \frac{\alpha}{2}\sum_{n=1}^{l-1}(l-n)na_{l-n}a_n - \alpha\sum_{n=1}^{N-l}(l+n)na_{l+n}\overline{a}_n \tag{4.27}$$

where $2 \leq l \leq N-1$ and

$$\dot{a}_1 = (\alpha - 4)a_1 - \alpha\sum_{n=2}^{N}(1-n)n\overline{a}_{n-1}a_n \tag{4.28}$$

and

$$\dot{a}_N = N^2(\alpha - 4N^2)a_N + \frac{\alpha}{2}\sum_{n=1}^{N}(N-n)na_{N-n}a_n \tag{4.29}$$

4.4.3 KL-Galerkin

The Fourier basis is a general basis for periodic functions. However, it may not provide the minimal number of equations D for a given partial differential equation. For the KS equation, the Fourier basis is in fact an optimal basis for parameter values α for which the solutions to the KS equation are translationally invariant ($u(x,t)$ is a solution implies $u(x+\beta,t)$ is also a solution). For values of α for which the solutions are not translationally invariant we may sensibly look for an optimal basis consisting of KL eigenvectors $\{\phi^{(n)}(x)\}$ which are not sinusoidal.

The derivation of the system of ODEs based on the KL basis is directly analogous to the Fourier Galerkin projection. The new decomposition takes the form

$$u(x,t) = \sum_{n=1}^{\infty} a_n(t)\phi^{(n)}(x). \tag{4.30}$$

Thus

$$\dot{a}_l = -4\sum_n a_n(\phi^{(l)}, \phi^{(n)}_{xxxx}) - \alpha\sum_n a_n(\phi^{(l)}, \phi^{(n)}_{xx}) - \frac{\alpha}{2}\sum_{mn}(\phi^{(l)}, \phi^{(m)}_x\phi^{(n)}_x)a_m a_n. \tag{4.31}$$

Now representing the KL eigenfunctions themselves in terms of their Fourier coefficients, i.e., letting

$$\phi^{(l)} = \sum_{k=-N}^{N} \alpha_k^{(l)} e^{ikx} \tag{4.32}$$

then an M term KL expansion takes the form

$$\dot{a}_l = \sum_{n=1}^{M} q_{ln} a_n + \sum_{m,n=1}^{M} p_{lmn} a_m a_n, \tag{4.33}$$

where

$$q_{ln} = \sum_{k=-N}^{N} (-4k^4 - k^2) \alpha_k^{(l)} \overline{\alpha}_k^{(n)} \tag{4.34}$$

and

$$p_{lmn} = -\alpha \sum_{k,k'=-N}^{N} k k' \alpha_{k+k'}^{(l)} \overline{\alpha}_k^{(m)} \overline{\alpha}_{k'}^{(n)} \tag{4.35}$$

with the restriction that $\mid k + k' \mid \leq N$.

Example 4.6. *The KS equation with* $\alpha = 84.25$. This procedure was carried out for the KS equation at a value of α for which the solutions were not translationally invariant. Three modes, shown in Figure 4.4, dominate the energy in the covariance matrix spectrum. The

A 3-dimensional simulation of the KS equation based on these modes is given below. Note that the truncated system requires a cubic stablizing term. For a comparison of the original solution with the reduced solution see Figure 4.5. Further details may be found in [43].

Fig. 4.4 The eigenvectors of the KS equation for $\alpha = 84$.

$$
\begin{aligned}
\dot{a}_1 &= (4.0001\alpha - 64.0110)a_1 + (0.0395 - 0.0021\alpha)a_2 \\
&\quad +(0.0013\alpha - 0.0256)a_3 + \alpha(0.0003a_1^2 + 0.0039a_1a_2 \\
&\quad +0.0109a_1a_3 - 0.2363a_2^2 - 0.3986a_2a_3 + 0.2363a_3^2) \\
&\quad -2.3a_1^3
\end{aligned}
$$

$$
\begin{aligned}
\dot{a}_2 &= (0.0395 - 0.0021\alpha)a_1 + (1.0036\alpha - 4.1637)a_2 \\
&\quad +(0.0001\alpha - 0.0052)a_3 + \alpha(0.0010a_1^2 + 1.1142a_1a_2 \\
&\quad +0.9644a_1a_3 + 0.0024a_2^2 - 0.0031a_2a_3 + 0.0034a_3^2) \\
&\quad -2.3a_2^3
\end{aligned}
$$

$$
\begin{aligned}
\dot{a}_3 &= (0.0013\alpha - 0.0256)a_1 + (0.0001\alpha - 0.0052)a_2 \\
&\quad +(1.0038\alpha - 4.1734)a_3 + \alpha\,(0.0016a_1^2 + 0.9595a_1a_2 \\
&\quad -1.1142a_1a_3 - 0.0051a_2^2 - 0.0020a_2a_3 + 0.0012a_3^2) \\
&\quad -2.3a_3^3
\end{aligned}
$$

4.4.4 A Weighted Sobolev Norm

The KL procedure was developed as an optimal procedure for empirically representing point values of a given data set in the mean-square sense. Note, however, that the numerical simulation of a PDE requires more than the optimal approximation of the data $u(x,t)$. It is also necessary to approximate

Fig. 4.5 Left: original 20-dimensional simulation. Right: reduced 3-dimensional simulation.

the spatial partial derivates of $u(x,t)$ at every time step. For instance, in the Kuramoto-Sivashinsky (KS) equation the solutions are highly sensitive to the dissipative term u_{xxxx} and the degree to which this is accurately approximated will greatly affect the results of the numerical simulation.

We observe that a basis which optimally represents $u(x,t)$ will not, in general, be optimal for the spatial derivatives $u_x, u_{xx}, u_{xxx}, u_{xxxx}$. The optimization criterion may be shifted to the derivatives employing the modified variational principle

$$\epsilon = \min\langle w_0 \parallel u - u^N \parallel^2 + w_1 \parallel u_x - u_x^N \parallel^2 + w_2 \parallel u_{xx} - u_{xx}^N \parallel^2 + \ldots \rangle \tag{4.36}$$

where the w_i are adjustable non-negative weights. Alternatively, we choose to maximize

$$\lambda_1 = \langle w_0(\phi^{(1)}, u)^2 + w_1(\phi^{(1)}, u_x)^2 + w_2(\phi^{(1)}, u_{xx})^2 + \ldots \rangle \tag{4.37}$$

subject to $(\phi^{(1)}, \phi^{(1)}) = 1$. The remaining eigenfunctions may be found by proceeding inductively as before.

Now the variational equation leads to the eigenfunction problem

$$\int K_S(x,y)\phi(y)dy = \lambda\phi(x) \tag{4.38}$$

where we have the modified kernel

$$K_S = \langle w_0 u \otimes u + w_1 u_x \otimes u_x + w_2 u_{xx} \otimes u_{xx} + \ldots \rangle \tag{4.39}$$

and $\otimes$ denotes an outer product.

How many derivatives we choose to include in the variational equation will depend on the form of the operator and on whether the resulting integral equation has solutions. Let us define the generalized Sobolev K-L kernel $K_S^{(j)}$ of order j

$$K_S^{(j)}(x,y) = \sum_{k=0}^{j} w_k \langle u^{(k)} \otimes u^{(k)} \rangle \tag{4.40}$$

where $u^{(k)}$ denotes the kth partial derivative. Details of the implementation of this optimality criterion are provided in [42].

4.4.5 Further Discussion

The procedure outlined above for constructing low-dimensional systems from infinite dimensional PDEs deserves further discussion. As a general approach it has enormous potential. It is worth addressing some of the obvious criticisms of this approach.

> *In order to construct the empirical basis the KL procedure requires data representing solutions. In essence, you have to first compute the solution to provide a method for computing the solution.*

This remark is certainly true, the KL basis is empirical and requires that data associated with the solution be known. The strength of the KL procedure is exactly its ability to exploit the *known* geometry of phase space. The drawback of the method, i.e., that solution sets be available is mitigated by several factors. In many instances, detailed numerical solutions are in fact available. Alternatively, experimental data sets may be available for basing the computation of the covariance matrix.

> *The reduced KL model is of limited validity in parameter space.*

A model built using the KL eigenvectors computed using data collected at a fixed parameter value α_0 may indeed have a limited range of validity $\alpha \in (\alpha_0 - a, \alpha_0 + a)$. This depends on the manner in which the phase space of the PDE changes with its parameter. In some instances the KL basis is actually quite robust to changes in parameter; see **??** for an example. This situtation is a fact of life for this approached to reduced systems. It depends on having an accurate representation for the geometry of the phase space of the solutions. If the geometry is not accurate then more data needs to be collected. Adaptive bases methods appear to be a very promising way around this problem.

> *The KL procedure is linear and for this reason may not be able to determine an optimal system of equations.*

The KL procedure is limited by the fact that the transformation is restricted to by orthogonal. Situations often arise where there is no *rotation* of phase space to reveal the low-dimensional dynamics. For example, consider the instance where the solutions for a particular parameter in the PDE lie on

a closed curve in a high-dimensional space. If the solution has its energy distributed in every direction then the KL procedure will not be able to reduce the system at all. However, geometrically we expect a 2-dimensional solution should be sufficient. This is due to the fact that the procedures described here produce best bases which encapsulate the appropriate regions of phase space for constructing simulations of reduced dimension. An alternative is to *parameterize* the solutions. See Chapter 9 for a discussion of these methods.

4.5 IMPLEMENTATION WITH MISSING DATA

Now we turn to the problem of using the KL procedure on data sets which have gaps, or missing components. The algorithm presented here is due to Everson and Sirovich [19]. Our development follows [19], although here we simplify the setting of the presentation using only discrete vector spaces, rather than function spaces. We distinguish this extension of the KL procedure for *gappy data*, using the terminology of [19], from the case of *noisy data* which is developed in the next section.

4.5.1 Estimating Missing Data

Let $\mathbf{x} \in \mathbb{R}^N$ be a vector which possesses a reduced expansion in terms of the KL basis as

$$\mathbf{x} \approx \mathbf{x}_D = \sum_{n=1}^{D} a_n \mathbf{u}^{(n)}$$

It follows that only D points of information are required to reproduce the original vector. Consider now an incomplete, or gappy, copy $\tilde{\mathbf{x}}$ of the original vector $\mathbf{x}$. This may be expressed

$$\tilde{x}_i = \begin{cases} x_i & m_i = 1, \\ 0 & m_i = 0 \end{cases} \tag{4.41}$$

where the vector $\mathbf{m} \in \mathbb{R}^N$ is in indicator vector, or *mask*, which identifies the indices of the missing data. We will also write this incomplete vector as

$$\tilde{\mathbf{x}} = \mathbf{m}.\mathbf{x}$$

where the i'th component of the product notation $(\mathbf{m}.\mathbf{x})_i = m_i x_i$ represents pointwise multiplication.

Given a vector $\mathbf{x}$ which is an element of an ensemble of intrinsically low-dimension, it may be possible to replace, or at least estimate, the missing entries. If the ambient dimension in which the vector resides is large, specifically if $D << N$, it is plausible that this repair may be possible even if a significant number of the entries of $\mathbf{x}$ are missing. The repaired vector $\tilde{\mathbf{x}}_D$ is

defined as an approximation to $\mathbf{x}$ as

$$\tilde{\mathbf{x}} \approx \tilde{\mathbf{x}}_D = \sum_{n=1}^{D} \tilde{a}_n \mathbf{u}^{(n)} \tag{4.42}$$

Note that, by virtue of the eigenvectors $\{\mathbf{u}^{(n)}\}$ being fully intact, the reconstruction $\tilde{\mathbf{x}}_D$ has no missing entries.

Now the problem is to find the $\{\tilde{a}_n\}$ such that

$$E = \|\tilde{\mathbf{x}} - \tilde{\mathbf{x}}_D\|_{\mathbf{m}}^2 \tag{4.43}$$

is a minimum where the norm is defined only on the data which is not missing, i.e.,

$$\|\mathbf{x}\|_{\mathbf{m}}^2 = (\mathbf{x}, \mathbf{x})_{\mathbf{m}} = (\mathbf{m}.\mathbf{x}, \mathbf{m}.\mathbf{x})$$

With this norm it follows that the coefficients $\{\tilde{a}_n\}$ are estimated based on the available data only. Using this definition of the (gappy) inner product

$$\begin{aligned} E &= (\tilde{\mathbf{x}} - \sum_{n=1}^{D} \tilde{a}_n \mathbf{u}^{(n)}, \tilde{\mathbf{x}} - \sum_{m=1}^{D} \tilde{a}_m \mathbf{u}^{(m)})_{\mathbf{m}} \\ &= (\tilde{\mathbf{x}}, \tilde{\mathbf{x}})_{\mathbf{m}} - 2(\tilde{\mathbf{x}}, \sum_{n=1}^{D} \tilde{a}_n \mathbf{u}^{(n)})_{\mathbf{m}} + (\sum_{n=1}^{D} \tilde{a}_n \mathbf{u}^{(n)}, \sum_{m=1}^{D} \tilde{a}_m \mathbf{u}^{(m)})_{\mathbf{m}} \\ &= \|\tilde{\mathbf{x}}\|_{\mathbf{m}}^2 - 2 \sum_{n=1}^{D} \tilde{a}_n (\tilde{\mathbf{x}}, \mathbf{u}^{(n)})_{\mathbf{m}} + \sum_{m,n=1}^{D} \tilde{a}_m \tilde{a}_n (\mathbf{u}^{(m)}, \mathbf{u}^{(n)})_{\mathbf{m}} \end{aligned}$$

Note that the eigenvectors $\{\mathbf{u}^{(m)}\}$ are no longer orthogonal on the gappy inner product.

Differentiating the error term E w.r.t. the k'th coefficient gives

$$\frac{\partial E}{\partial \tilde{a}_k} = 0 - 2(\tilde{\mathbf{x}}, \mathbf{u}^{(k)})_{\mathbf{m}} + 2 \sum_{m=1}^{D} \tilde{a}_m (\mathbf{u}^{(m)}, \mathbf{u}^{(k)})_{\mathbf{m}} = 0$$

from which it follows that

$$\sum_{m=1}^{D} \tilde{a}_m (\mathbf{u}^{(m)}, \mathbf{u}^{(k)})_{\mathbf{m}} = (\tilde{\mathbf{x}}, \mathbf{u}^{(k)})_{\mathbf{m}}$$

This may be rewritten in the form of a linear system

$$M\tilde{\mathbf{a}} = \mathbf{f} \tag{4.44}$$

where

$$M_{ij} = (\mathbf{u}^{(i)}, \mathbf{u}^{(j)})_{\mathbf{m}}$$

and

$$f_i = (\tilde{\mathbf{x}}, \mathbf{u}^{(i)})_{\mathbf{m}}$$

4.5.2 Estimating a KL Basis with Missing Data

In the previous section we examined the question of estimating data missing from an observation. The procedure required a KL basis derived from a training set with no gaps. These ideas may now be applied to constructing a KL basis where only incomplete data sets are available. The procedure presented here was proposed in [19]. It is based on an iterative process which successively *repairs* the gappy data and improves the estimate for the associated KL basis.

The data set may be modeled by associating with each pattern, a mask $\mathbf{m}^{(\mu)}$ of indices indicating which data is available and which components are missing. Each pattern with incomplete data may now be written

$$\tilde{\mathbf{x}}^{(\mu)} = \mathbf{m}^{(\mu)}.\mathbf{x}^{(\mu)}$$

The ensemble average of the incomplete patterns is now

$$(\langle\tilde{\mathbf{x}}\rangle)_i = \frac{1}{P_i}\sum_{\mu=1}^{P}\tilde{x}_i^{(\mu)}$$

where

$$P_i = \sum_{\mu=1}^{P} m_i^{(\mu)}$$

Once this ensemble average has been determined from the gappy data, the first stage of the ensemble *repair* procedure may be executed. This repair is done by replacing the missing data with the pointwise mean of the existing data. Specifically, the first stage of the repair process is then

$$x_i^{(\mu)}(0) = \begin{cases} \tilde{x}_i^{(\mu)} & m_i^{(\mu)} = 1, \\ (\langle\tilde{\mathbf{x}}\rangle)_i & m_i^{(\mu)} = 0 \end{cases} \tag{4.45}$$

The improved ensemble $\{\mathbf{x}^{(\mu)}(0)\}$ may be used to construct the first estimate of, or initialize, the KL basis, which we denote

$$\{\mathbf{u}^{(j)}(0)\}_{j=1}^{P}$$

Now, given an initial estimate for the KL basis vectors, an improved approximation may be obtined using the procedure of the previous section.

Specifically, given the gappy pattern vector $\tilde{\mathbf{x}}^{(\mu)} = \mathbf{m}^{(\mu)}\mathbf{x}^{(\mu)}$ we may improve our estimate of $\mathbf{x}$ given by Equation (4.45) by using the first estimate of the KL basis. The improved estimate may be written

$$\tilde{\mathbf{x}}_D^{(\mu)}(1) = \sum_{m=1}^{D}\tilde{a}_m^{(\mu)}(1)\mathbf{u}^{(m)}(0)$$

where the $\{\tilde{a}_m^{(\mu)}(1)\}$ are the solutions to

$$M^{(\mu)}(0)\tilde{\mathbf{a}}^{(\mu)}(1) = \mathbf{f}^{(\mu)}(0)$$

where

$$M_{ij}^{(\mu)}(0) = (\mathbf{u}^{(i)}(0), \mathbf{u}^{(j)}(0))_{\mathbf{m}^{(\mu)}}$$

and

$$f_k^{(\mu)}(0) = (\tilde{\mathbf{x}}^{(\mu)}(0), \mathbf{u}^{(k)}(0))_{\mathbf{m}^{(\mu)}}$$

The second iteration of the repair procedure uses the improved estimate for the gappy data ensemble

$$x_i^{(\mu)}(1) = \begin{cases} \tilde{x}_i^{(\mu)} & m_i^{(\mu)} = 1, \\ (\tilde{\mathbf{x}}_D^{(\mu)}(1))_i & m_i^{(\mu)} = 0 \end{cases} \tag{4.46}$$

In summary, a set of KL basis vectors may be estimated from gappy data as follows:

KL Procedure for Gappy Data

1. Initialize the missing data with the ensemble average.
2. Using the completed data compute the first iteration of the KL basis.
3. Re-estimate the ensemble using the gappy approximation from the KL basis.
4. Re-compute the KL basis.

The last two steps of this process are repeated until the KL basis is deemed to have converged in a satisfactory manner. In particular, we expect the the sequence of repairs to approach the actual data

$$\mathbf{x}^{(\mu)}(n) \to \mathbf{x}^{(\mu)}$$

and consequently, the sequence of estimated eigenvectors to approach the actual eigenvectors

$$\mathbf{u}^{(i)}(n) \to \mathbf{u}^{(i)}.$$

It is natural to end the iteration when the updates provide little change and it is concluded that no further progress is being made.

4.6 APPLICATION TO NOISY DATA

Now we turn to the case where the patterns have added noise, i.e.,

$$\mathbf{x}^{(\mu)} = \mathbf{s}^{(\mu)} + \mathbf{n}^{(\mu)}$$

In the general situation, the signal $\mathbf{s}^{(\mu)}$ is not observable, but the signal with the added noise component $\mathbf{n}^{(\mu)}$ is. A computation of the best basis in this

situation, using the ensemble averaged covariance matrix $C_{\mathbf{x}} = \langle \mathbf{x}\mathbf{x}^T \rangle$ does not, in general, provide a good separation of the signal and the noise.

The simplest case is that of *white* noise which is assumed to have zero mean, be uncorrelated with the signal and have a covariance matrix of the form αI where α is the variance of the noise and I is the identity matrix. In this instance, the covariance matrix of the signal may be decomposed

$$\begin{aligned} C_{\mathbf{x}} &= \langle (\mathbf{s}+\mathbf{n})(\mathbf{s}+\mathbf{n})^T \rangle \\ &= \langle \mathbf{s}\mathbf{s}^T \rangle + \langle \mathbf{n}\mathbf{n}^T \rangle \\ &= C_{\mathbf{s}} + \alpha I \end{aligned}$$

where $C_{\mathbf{s}} = \langle \mathbf{s}\mathbf{s}^T \rangle$. The eigenvectors of $C_{\mathbf{x}}$ are the same as $C_{\mathbf{s}}$ and the eigenvalues are all shifted upwards by the variance of the noise α, leaving the differences of the eigenvalues preserved [22].

The general situtation is more complicated and the noise does indeed change the eigenvectors. Now the signal and noise may be separated in an optimal sense by defining an appropriate variational principle. To start, write down expansions for the noise

$$\mathbf{n}^{(\mu)} = \sum_i b_i^{(\mu)} \psi^{(i)}$$

and the observed variable

$$\mathbf{x}^{(\mu)} = \sum_i a_i^{(\mu)} \psi^{(i)}$$

Following [24], a basis for separating the signal from the noise may be obtained by determining (ideally) an o.n. basis $\{\psi^{(i)}\}$ such that

$$\frac{\langle b_i^2 \rangle}{\langle a_i^2 \rangle} = \text{maximum}$$

for each direction i.

From our experience with the derivations in the previous sections it is natural to consider first maximizing

$$D(\psi^{(1)}) = \frac{\langle b_1^2 \rangle}{\langle a_1^2 \rangle}$$

Before differentiating, we rewrite D as

$$D(\psi^{(1)}) = \frac{(\boldsymbol{\psi}^{(1)}, \langle \mathbf{n}\mathbf{n}^T \rangle \boldsymbol{\psi}^{(1)})}{(\boldsymbol{\psi}^{(1)}, \langle \mathbf{x}\mathbf{x}^T \rangle \boldsymbol{\psi}^{(1)})}$$

Observe the appearance of the ensemble averaged noise covariance matrix which we write $C_{\mathbf{n}} = \langle \mathbf{n}\mathbf{n}^T \rangle$, in addition to the usual matrix $C_{\mathbf{x}} = \langle \mathbf{x}\mathbf{x}^T \rangle$.

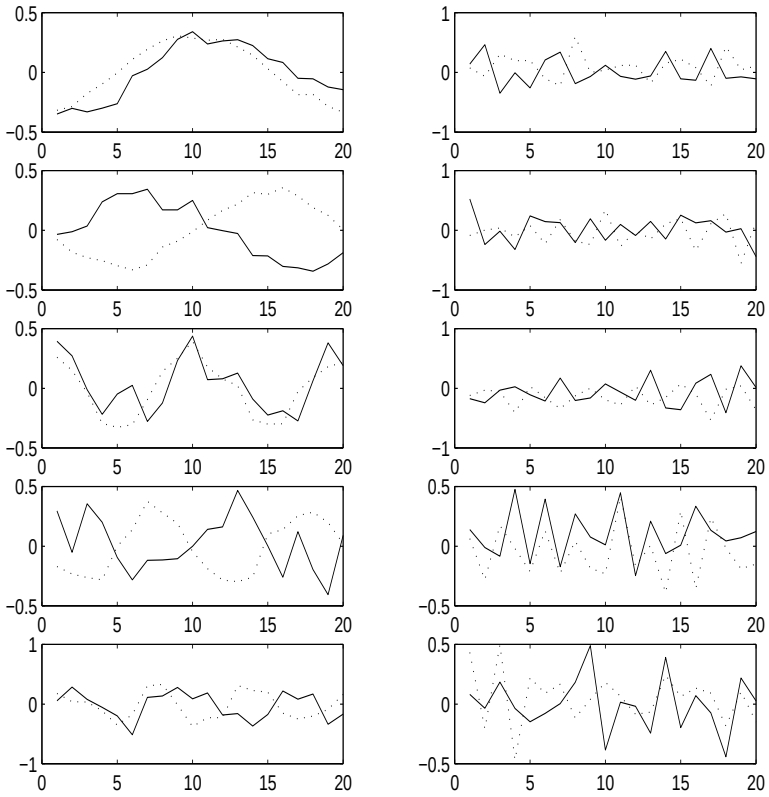

Fig. 4.6 Left column top to bottom: The eigenvectors corresponding to the largest eigenvalues of $C_{\mathbf{x}}$ are represented by solid lines and the eigenvectors corresponding to the smallest eigenvalues of $C_{\mathbf{x}}^{-1}C_{\mathbf{n}}$ are represented by dotted lines. Right column top to bottom: The eigenvectors corresponding to the smallest eigenvalues of $C_{\mathbf{x}}$ and the largest eigenvalues of $C_{\mathbf{x}}^{-1}C_{\mathbf{n}}$.

Now we seek to maximize

$$D(\psi^{(1)}) = \frac{(\psi^{(1)}, C_{\mathbf{n}}\psi^{(1)})}{(\psi^{(1)}, C_{\mathbf{x}}\psi^{(1)})} \tag{4.47}$$

Note that it is not necessary to require the auxiliary condition that $(\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(1)}) = 1$ since $D(\psi^{(1)})$ is now bounded. Differentiating Equation 4.47 w.r.t. $\psi^{(1)}$ gives

$$\frac{\partial D}{\partial \boldsymbol{\psi}^{(1)}} = 2C_{\mathbf{n}}\boldsymbol{\psi}^{(1)}(\boldsymbol{\psi}^{(1)}, C_{\mathbf{x}}\boldsymbol{\psi}^{(1)}) - 2C_{\mathbf{x}}\boldsymbol{\psi}^{(1)}(\boldsymbol{\psi}^{(1)}, C_{\mathbf{n}}\boldsymbol{\psi}^{(1)}) = 0 \qquad (4.48)$$

This can be rewritten as

$$C_{\mathbf{n}}\boldsymbol{\psi}^{(1)} = \mu C_{\mathbf{x}}\boldsymbol{\psi}^{(1)} \qquad (4.49)$$

where

$$\mu = \frac{(\boldsymbol{\psi}^{(1)}, C_{\mathbf{n}}\boldsymbol{\psi}^{(1)})}{(\boldsymbol{\psi}^{(1)}, C_{\mathbf{x}}\boldsymbol{\psi}^{(1)})}$$

This is a *symmetric definite generalized eigenproblem*, see [?].

Compare with the standard eigenvector problem

$$B\boldsymbol{\psi}^{(1)} = \lambda\boldsymbol{\psi}^{(1)}$$

If we take $B = C_{\mathbf{x}}^{-1}C_{\mathbf{n}}$ we need to verify that the μ is in fact an eigenvalue of B. We will show that if

$$C_{\mathbf{x}}^{-1}C_{\mathbf{n}}\boldsymbol{\psi}^{(1)} = \lambda\boldsymbol{\psi}^{(1)}$$

then λ has the form given for μ. This is readily seen to be true by projecting the eigenvector equation onto $\boldsymbol{\psi}^{(1)}$, i.e.,

$$(\boldsymbol{\psi}^{(1)}, C_{\mathbf{n}}\boldsymbol{\psi}^{(1)}) = \lambda(\boldsymbol{\psi}^{(1)}, C_{\mathbf{x}}\boldsymbol{\psi}^{(1)})$$

from which we see that λ has exactly the desired form.

The remaining optimal basis vectors may found using a similiar approach. Note, however, that in this case the eigenvector problem is no longer symmetric in general. Using this approach, the ordering of the eigenvectors is reversed. The noise resides in the largest eigenvalues while the signal is spanned by the basis vectors with the smallest eigenvalues.

To successfully implement this method it is necessary to estimate the ensemble averaged covariance matrix of the noise $C_{\mathbf{n}}$. Note that although, following [24], the method was introduced in the context of eliminating noise, it is useful for separating signals in general when the ensemble averaged covariance matrices for each component are available.

Example 4.7. *Fourier modes corrupted with known noise.* As a simple example of the procedure we generate data according to

$$u(x,t) = \sum_{k=1}^{N} \frac{1}{k} \sin k(t-k) + \eta(x,t)$$

where $\eta(x,t)$ consists of Gaussian noise with mean zero and variance four. Taking $N_x = 20$ and $N_t = 5000$ the data were discretized according to $x_i = 2\pi i/N_x$, $i = 1, \ldots, N_x$ and $t_j = 2\pi j/N_t$, $j = 1, \ldots, N_t$. In addition N was taken as 40. In the limit, as $N_t \to \infty$, the matrices $C_{\mathbf{n}}$ and $C_{\mathbf{n}}^{-1}$ approach the identity. Hence, given infinite samples, this approach would be the same as standard KL.

4.7 THE LOCAL KARHUNEN-LOÈVE EXPANSION

The Karhunen-Loève procedure described in the previous sections provides a *global* representation for a data set. It is not unusual, however, for data to reside on a nonlinear surface, or *manifold*, which appears locally to possess a coordinate system of reduced dimension. In this setting, the optimal global basis may require an encapsulating dimension which far exceeds the local, or topological dimension of the data. In this section we discuss a local representation which overcomes this deficiency.

The main points to be addressed in this section are how

- a data surface may be locally approximated by a tangent space
- the tangent space may be computed via a local KL procedure
- the local dimension may be estimated using a scaling argument
- the procedure may be automated using a local whitening transformation

In addition, we shall see in Chapter 9 Section 9.11 that data may be locally parameterized by the coordinates of the local basis.

4.7.1 Global KL Procedure on Closed Curves

Before presenting a general approach for computing local bases it will be useful to examine the application of the KL procedure to several sample problems. The problems reveal a pattern which may be exploited to develop an algorithm.

Example 4.8. *Eigenvectors of a Circle.* To begin, we consider the calculation of the KL eigenvectors for data on the unit circle S^1. The circle may be parameterized by the angle θ

$$(x(\theta), y(\theta) = (\cos\theta, \sin\theta)$$

Given any point is uniquely determined by a single parameter, the set of points is one-dimensional. However, given the symmetry of the circle, we anticipate that a global KL basis will be two-dimensional.

Since the data is defined continuously on the circle, we now employ the integral definition of the ensemble average

$$\langle x \rangle = \frac{1}{2\pi} \int_0^{2\pi} x(\theta) d\theta$$

It is easily verified that the mean value on the circle is the origin since $\langle x \rangle = \langle y \rangle = 0$. The ensemble averaged covariance matrix $\mathbf{C}$ is then given by

$$\mathbf{C} = \begin{pmatrix} \langle x^2 \rangle & \langle xy \rangle \\ \langle xy \rangle & \langle y^2 \rangle \end{pmatrix}$$

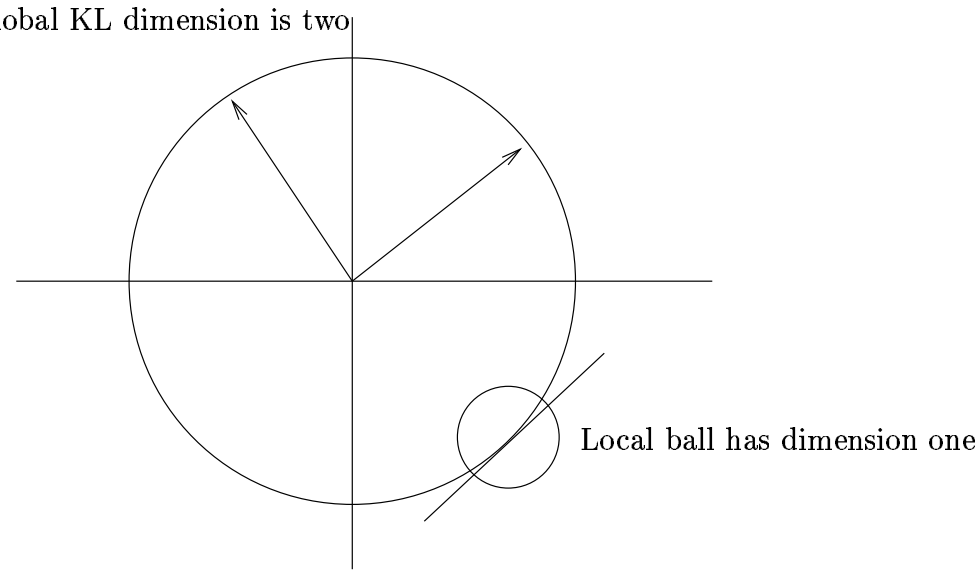

Fig. 4.7 The circle appears to be two-dimensional globally. In a local ball, the KL procedure identifies data on the circle as one-dimensional.

The components of the matrix are readily computed as

$$\langle x^2 \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos^2 \theta d\theta = \frac{1}{2},$$

$$\langle y^2 \rangle = \frac{1}{2\pi} \int_0^{2\pi} \sin^2 \theta d\theta = \frac{1}{2},$$

$$\langle xy \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos\theta \sin\theta d\theta = 0.$$

Thus, we see that the matrix $\mathbf{C}$ is *already* diagonalized from which we conclude the standard coordinate system provides uncorrelated axes. Thus, with

$$\mathbf{C} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

we see that all vectors in $\mathbb{R}^2$ are eigenvectors and any orthogonal pair qualifies as an optimal KL basis. In other words, every coordinate system has uncorrelated data.

Example 4.9. *Eigenvectors of an Ellipse.* Now we compute the optimal eigenvectors and the associated eigenvalues for the family of ellipses rotated

in the plane by an angle α. The parameterized equations are given as

$$x(\theta) = a\cos\alpha\cos\theta - b\sin\alpha\sin\theta$$
$$y(\theta) = a\sin\alpha\cos\theta + b\cos\alpha\sin\theta$$

It is again easy to verify that $\langle x\rangle = \langle y\rangle = 0$. The remaining components of the covariance matrix are

$$\begin{aligned}
\langle x^2\rangle &= \frac{1}{2\pi}\int_0^{2\pi}(a\cos\alpha\cos\theta - b\sin\alpha\sin\theta)^2 d\theta \\
&= \frac{1}{2}(a^2\cos^2\alpha + b^2\sin^2\alpha) \\
\langle y^2\rangle &= \frac{1}{2\pi}\int_0^{2\pi}(a\sin\alpha\cos\theta + b\cos\alpha\sin\theta)^2 d\theta \\
&= \frac{1}{2}(a^2\sin^2\alpha + b^2\cos^2\alpha) \\
\langle xy\rangle &= \frac{1}{2\pi}\int_0^{2\pi}(a\cos\alpha\cos\theta - b\sin\alpha\sin\theta)(a\sin\alpha\cos\theta + b\cos\alpha\sin\theta) d\theta \\
&= (a^2 - b^2)\sin(2\alpha)
\end{aligned}$$

We may facilitate our computations by rotating coordinates in a manner whcih diagonalizes the covariance. Since we anticipate that the eigenvectors will in fact be scalar multiples of $\mathbf{w}_1 = (\cos\alpha, \sin\alpha)^T$ and $\mathbf{w}_2 = (-\sin\alpha, \cos\alpha)^T$, we attempt to diagonalize $\mathbf{C}$ by computing $\mathbf{C}' = \mathbf{Q}\mathbf{C}\mathbf{Q}^T$ where

$$\mathbf{Q}^T = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

is the matrix which has columns $\mathbf{w}_1, \mathbf{w}_2$. In fact, this transformation is equivalent to rotating the axes of the ellipse by the angle $-\alpha$ to align them with the standard basis. Carrying out this transformation we find $\mathbf{C}'$ is given as

$$\mathbf{C}' = \begin{pmatrix} a^2/2 & 0 \\ 0 & b^2/2 \end{pmatrix}$$

Hence $\lambda_1 = a^2/2$ contains $100 \times a^2/(a^2 + b^2)$ percent of the energy. We see that for $a = b$, i.e., the case of the circle, we obtain the same result as before.

4.7.2 The Local Approach

In the previous section we saw two examples of analytically defined curves for which it was possible to compute local KL eigenvectors and eigenvalues. These curves appear locally like straight lines (as the local region shrinks to radius zero). It is a feature of k-dimensional manifolds that the appear locally like $\mathbb{R}^k$ regardless of the dimension of the ambient space. The local approach to be described here exploits the locally Euclidean nature of data on a manifold.

Before exploring k dimensional manifolds we examine one-dimensional manifolds, i.e., smooth curves, and outline a local KL procedure for representing these curves.

We begin by returning to our circle example. We propose to construct a local basis about the point $(\cos\alpha, \sin\alpha)$. Now the data is parameterized as

$$(x, y) = (\cos\theta - \cos\alpha, \sin\theta - \sin\alpha).$$

The components of the ensemble averaged covariance matrix are given by

$$C_{11} = \langle x^2 \rangle = \frac{2\pi}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} (\cos\theta - \cos\alpha)^2 d\theta$$

$$C_{22} = \langle y^2 \rangle = \frac{2\pi}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} (\sin\theta - \sin\alpha)^2 d\theta$$

$$C_{12} = \langle xy \rangle = \frac{2\pi}{2\epsilon} \int_{\alpha-\epsilon}^{\alpha+\epsilon} (\cos\theta - \cos\alpha)(\sin\theta - \sin\alpha) d\theta$$

where we are integrating locally over an arc of length $2\epsilon/2\pi$. Note that the data is mean subtraced only in the limit as $\epsilon \to \infty$ since for $\epsilon > 0$ the actual mean of the data does not lie on the circle. These integrals evaluate to

$$C_{11} = \frac{\pi}{\epsilon}(\epsilon - \frac{1}{2}\sin 2\epsilon + (2\epsilon + \sin 2\epsilon - 4\sin\epsilon)\cos^2\alpha)$$

$$C_{22} = \frac{\pi}{\epsilon}(3\epsilon + \frac{1}{2}\sin 2\epsilon - 4\sin\epsilon - (2\epsilon + \sin 2\epsilon - 4\sin\epsilon)\cos^2\alpha)$$

$$C_{12} = C_{21} = \frac{\pi}{\epsilon}(\frac{1}{2}\sin 2\epsilon - 2\sin\epsilon + \epsilon)$$

Since we anticipate that the eigenvectors will be scalar multiples of $\mathbf{w}_1 = (\sin\alpha, -\cos\alpha)^T$ and $\mathbf{w}_2 = (\cos\alpha, \sin\alpha)^T$, we may attempt to diagonalize C by computing

$$C' = QCQ^T$$

where

$$Q^T = \begin{pmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{pmatrix}$$

Carrying out this transformation we find C' is given as

$$C'(\epsilon) = \begin{pmatrix} \frac{\pi}{\epsilon}(\epsilon - \frac{1}{2}\sin 2\epsilon) & 0 \\ 0 & \frac{\pi}{\epsilon}(3\epsilon + \frac{1}{2}\sin 2\epsilon - 4\sin\epsilon) \end{pmatrix}$$

Since $C'(\epsilon)$ is diagonal, we conclude that the columns of Q^T are in fact the KL eigenvectors with the associated eigenvalues being C'_{11} and C'_{22}. It is interesting to compare the growth of these eigenvalues with the local region

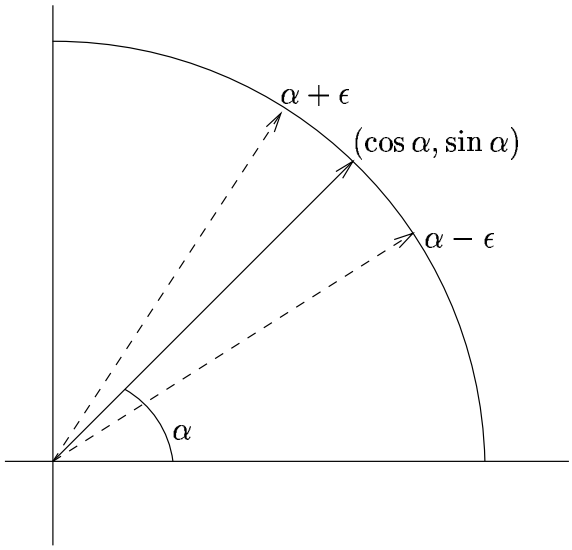

Fig. 4.8 Set-up for computing the KL basis locally on the circle. The data is taken to be the arc which extends from $\alpha - \epsilon \leq \theta \leq \alpha + \epsilon$

the size of which is a function of ϵ. Computing the Taylor expansions of the eigenvalues we obtain

$$C'_{11} = \sigma_1^2 = \frac{2\pi}{3}\epsilon^2 - \frac{2\pi}{15}\epsilon^4 + O(\epsilon^6)$$

$$C'_{22} = \sigma_2^2 = \frac{\pi}{8}\epsilon^4 - \frac{\pi}{84}\epsilon^6 + O(\epsilon^8)$$

where we consider the singular values $\sigma_i^2 = \lambda_i$ which are the square roots of the eigenvalues of $\mathbf{C}'$ [1].

Notice that the first singular value, the one corresponding to the eigenvector tangent to the data, grows linearly with ϵ. The second singular value, however, grows at a rate proportional to ϵ^2. We shall see this linear growth behavior is typical of the singular values corresponding to the eigenvectors that represesnt a flat approximation to the data, i.e., the tangent space.

It is interesting to compare the Taylor series expansion about the point $\mathbf{f}(\alpha) = (f_1(\alpha), f_2(\alpha)) = (\cos\alpha, \sin\alpha)$.

$$f_1(\alpha + \epsilon) = \cos\alpha - \epsilon\sin\alpha - \frac{\epsilon^2}{2}\cos\alpha + O(\epsilon^3)$$

[1] We say that $f(x) = O(g(x))$ as $x \to x_o$ if there exist positive constants K, δ such that $|f| \leq K|g|$ whenever $0 < |x - x_o| < \delta$.

$$f_2(\alpha+\epsilon) = \sin\alpha - \epsilon\cos\alpha - \frac{\epsilon^2}{2}\sin\alpha + O(\epsilon^3)$$

In vector notation,

$$\mathbf{f}(\alpha+\epsilon) = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} + \epsilon \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} + \frac{\epsilon^2}{2}\begin{pmatrix} -\cos\alpha \\ -\sin\alpha \end{pmatrix} + O(\epsilon^3)$$

Writing $\mathbf{w}_1 = (-\sin\alpha, \cos\alpha)^T$ and $\mathbf{w}_2 = (-\cos\alpha, -\sin\alpha)^T$ we have as the representation in the neigborhood of α

$$\delta\mathbf{f} = \epsilon\mathbf{w}_1 + \epsilon^2\mathbf{w}_2 + O(\epsilon^3).$$

These are exactly the eigenvectors computed above and that the coefficients scale as the singular values. However, in general, the Taylor series approximation will *not* coincide with the KL expansion.

Example 4.10. *An Example in* $\mathbb{R}^2$. Consider the quadratic curve

$$y = x^2.$$

What are the KL eigenvectors of the graph of the function (x, x^2) computed in a region about $(0,0)$? The diagonal terms of the covariance matrix are then

$$\langle x^2\rangle = \frac{1}{L}\int_{-\epsilon}^{\epsilon} x^2 dx$$

$$\langle y^2\rangle = \frac{1}{L}\int_{-\epsilon}^{\epsilon} x^4 dx$$

where L is the arclength of the curve. The diagonal terms are odd functions so

$$\langle xy\rangle = \frac{1}{L}\int_{-\epsilon}^{\epsilon} x^3 dx = 0$$

The arclength is defined as

$$L = \int_{-\epsilon}^{\epsilon} \sqrt{1+y'}dx$$

Noting that the Taylor series for

$$\sqrt{1+2x} = 1 + x - \frac{x^2}{2} + \frac{x^3}{2} - \frac{5}{8}x^4 + O(\epsilon^5)$$

it follows that

$$L = 2\epsilon - \frac{\epsilon^3}{3} + O(\epsilon^5)$$

Hence

$$\frac{1}{L} = \frac{1}{2\epsilon}(1 + O(\epsilon^2))$$

Now the diagonal entries may be evaluated as

$$\langle x^2 \rangle = \frac{\epsilon^2}{3}(1 + O(\epsilon^2))$$

and

$$\langle y^2 \rangle = \frac{\epsilon^4}{5}(1 + O(\epsilon^2))$$

Hence the covariance matrix is already diagonal and to leading order is given by

$$\mathbf{C} = \begin{pmatrix} \frac{\epsilon^2}{3} & 0 \\ 0 & \frac{\epsilon^4}{5} \end{pmatrix}$$

Unlike the circle example, the standard basis is in fact the unique (since the eigenvalues are distinct) KL basis.

Example 4.11. *Local KL on the sphere.* Recall that the surface of a sphere in $\mathbb{R}^3$ may be parameterized in terms of the position vector

$$\mathbf{x}(u, v) = x_1(u, v)\hat{\mathbf{i}} + x_2(u, v)\hat{\mathbf{j}} + x_3(u, v)\hat{\mathbf{k}}$$

where

$$x_1(u, v) = \cos u \cos v$$
$$x_2(u, v) = \sin u \cos v$$
$$x_3(u, v) = \sin v$$

and the domain T is the Cartesian product

$$T = [0, 2\pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$$

As the point $(u, v) \in T$ is varied the parameterization vector $(x_1(u, v), x_2(u, v), x_3(u, v))^T$ traces out the unit sphere centered at the origin. The covariance matrix for data over a patch S on the surface of the sphere is then defined

$$C_{ij} = \frac{1}{S} \int\int_S x_i x_j dS$$

where S is the surface area of the patch. It can be shown that the eigenvalues of C are given by

$$\lambda_1 = \lambda_2 = \frac{\epsilon^2}{4} - \frac{\epsilon^4}{24}$$

and

$$\lambda_3 = \frac{\epsilon^4}{12}$$

where ϵ is the radius of the ball enclosing the patch. The details of this calculation are left to the reader, see Exercise 4.16.

We see that the number of singular values (eigenvalues) which scale linearly with the size of local region is the same as the dimension of the tangent plane of the sphere. This is discussed further in the next section.

4.7.3 Local Representation and the Taylor Series

Now we return to the question of how a data surface is locally approximated in terms of its *tangent space*. Consider the observed data to be sampled from a vector function $\mathbf{f}$

$$\begin{aligned} \mathbf{f} : &\mathbb{R}^n \to \mathbb{R}^m \\ &\mathbf{x} \rightsquigarrow \mathbf{f}(\mathbf{x}) \end{aligned}$$

In particular, we are interested in the situation where only the range values of the samples $\{\mathbf{f}(\mathbf{x}^{(\mu)})\}$ are observed and that the associated domain values $\{\mathbf{x}^{(\mu)}\}$ are unknown. In addition, the dimension of the domain is also unknown.

Nonetheless, we may express this function as a Taylor series expansion about the point $\mathbf{x}_0$, i.e.,

$$\mathbf{f}(\mathbf{x}_o + \mathbf{h}) = \mathbf{f}(\mathbf{x}_o) + \mathbf{Df}(\mathbf{x}_o)\mathbf{h} + O(\mathbf{h}^2)$$

where $\mathbf{Df}(\mathbf{x}_o)$ is the Jacobian matrix of partial derivatives

$$\mathbf{Df}(\mathbf{x}_o) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

and $\mathbf{h} = (h_1, \ldots, h_n)$.

To emphasize its dependence on the distance $\mathbf{h}$, the expression for the local change in the function about the point $\mathbf{x}_o$ will be written $\delta\mathbf{f}(\mathbf{h}; \mathbf{x}_o) = \mathbf{f}(\mathbf{x}_o + \mathbf{h}) - \mathbf{f}(\mathbf{x}_o)$ The linear approximation to $\delta\mathbf{f}(\mathbf{h}; \mathbf{x}_o)$ is then

$$d\mathbf{f}(\mathbf{h}; \mathbf{x}_o) = \mathbf{Df}(\mathbf{x}_o)\mathbf{h} = h_1 \frac{\partial \mathbf{f}}{\partial x_1} + \cdots + h_n \frac{\partial \mathbf{f}}{\partial x_n}$$

With this notation, we may view the linear approximation $d\mathbf{f}(\mathbf{h}; \mathbf{x}_o)$ to $\delta\mathbf{f}(\mathbf{h}; \mathbf{x}_o)$ as a vector expansion. In fact, $d\mathbf{f}(\mathbf{h}; \mathbf{x}_o)$ lies in the vector space, centered at $\mathbf{x}_o$, spanned by the columns of $\mathbf{Df}(\mathbf{x}_o)$. The space spanned by these vectors, i.e., the range of the Jacobian about the point $\mathbf{x}_0$ is the tangent space $T_{\mathbf{x}_0}$. (Note that $d\mathbf{f}(0) = 0$ so the origin of $T_{\mathbf{x}_0}$ is the zero vector.) Therefore,

$$d\mathbf{f}(\mathbf{h}; \mathbf{x}_o) \in \text{span}\{\frac{\partial \mathbf{f}}{\partial x_1}, \ldots, \frac{\partial \mathbf{f}}{\partial x_n}\} = T_{\mathbf{x}_0}.$$

4.7.4 Computation of the Tangent Space

Now that the Jacobian has been identified as producing the best linear, or flat, approximation to a surface we seek to make the connection with the KL

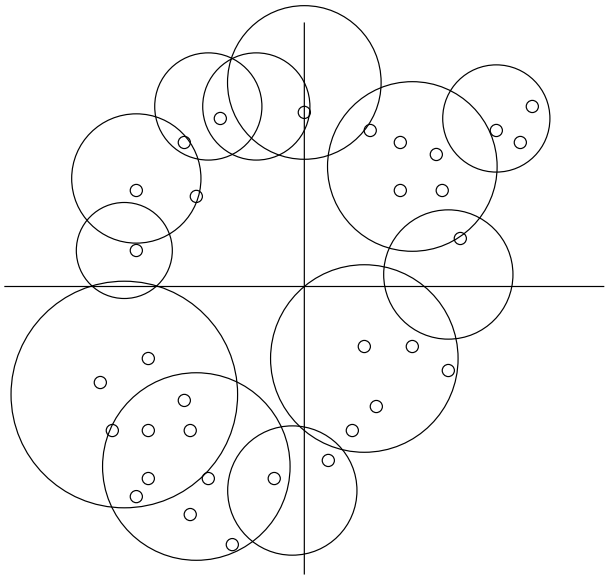

Fig. 4.9 A local covering of a data set in $\mathbb{R}^2$.

procedure. Imagine the limiting procedure of collapsing a ball about a point on the surface. As the ball becomes smaller and smaller, the curvature of the surface becomes less apparent. Indeed, by the time the ball is infinitessimally small, the surface looks flat. We would presume that the KL eigenvectors of this limiting ball would be span the Jacobian.

Following [22], it is possible to use the notion of KL dimension defined in Section 3.5 to estimate the spanning vectors of the Jacobian. However, this approach requires the estimation of an *ad-hoc* energy level. Another algorithm for computing the tangent space of a surface from locally sampled data may be based on the scaling of the singular values [9]. We now consider this approach in detail. The method is attractive given it requires no arbitrary energy criterion. It is based on the behavior of the singular values of a local region.

Definition 4.3. *An ϵ-neighborhood matrix $B_\epsilon(\mathbf{c})$ has columns made up of all the data vectors $\{x - \mathbf{c}\}$ in the open ball of radius ϵ centered at the point $\mathbf{c}$, i.e,*

$$\|\mathbf{x} - \mathbf{c}\| < \epsilon.$$

The scaling argument is based on computing the singular values of $\mathbf{B}_\epsilon(\mathbf{c})$ as a function of ϵ. Recall that the squared singular values are the eigenvalues of the covariance matrix

$$C(\epsilon) = B_\epsilon B_\epsilon^T$$

Proposition 4.3. *(Broomhead, Jones, King [9]). Consider the mapping*

$$\mathbf{f} : U \to X$$

about the point $\mathbf{x}_0 \in U$. *A basis for the tangent space* $T_{\mathbf{x}_0}$ *is provided by the singular values of the* ϵ*-neighborhood matrix* $B_\epsilon(\mathbf{x}_0)$ *that scale linearly with the radius* ϵ.

A proof of this proposition is provided in a general setting in [11]. A tangent space constructed in this manner is referred to as an ϵ-*tangent space* [9]. We now summarize the local KL algorithm proposed in [9].

The Local KL Algorithm

1. Define a set of (possibly overlapping) ϵ-matrix neighborhoods $\{B_\epsilon(\mathbf{c}^{(i)})\}$ which *cover* the data.
2. For each neighborhood, indexed by i, compute the singular values as a function of ϵ for $\epsilon \to 0$.
3. Identify the singular values which scale linearly with ϵ.
4. The associated eigenvectors form the local basis. The dimension of each basis is referred to as the *local KL dimension*.

Given sufficient data, the above algorithm can be quite effective for estimating local dimensions, see, e.g., [9, 11, 1].

Statistics versus Dimension. The graph in Figure 4.10 depicts the result of applying this algorithm to a neighborhood of uniformly sampled points on the surface of the unit sphere. The top graph of Figure 4.10 compares the numerically determined values with the analytical estimates (see Exercise 4.16) and we see that they are in agreement.

To show the importance of the manner in which the data is distributed in the neighborhood, we repeated the above experiment with deliberately non-uniformly sampled data. The singular values are shown in the middle graph of Figure 4.10; they still scale linearly with ϵ but now the slope the local SVD curve has changed.

From this simple example we conclude that the information contained in the singular values consists of two basic components: local geometry and local statistics. The geometry, i.e., the dimension of the tangent space, is determined by the scaling of the singular values while the slopes of the curve reflect the distribution of the data. In the next section we adapt the procedure to remove the local dependence on the statistics.

4.7.5 Local Whitening Transformation

The difficulty in applying the local KL scaling criterion rests in determining the exact nature, e.g., linear or quadratic, of the scaling of the singular values which is complicated due to the influence of the data statistics. In this section

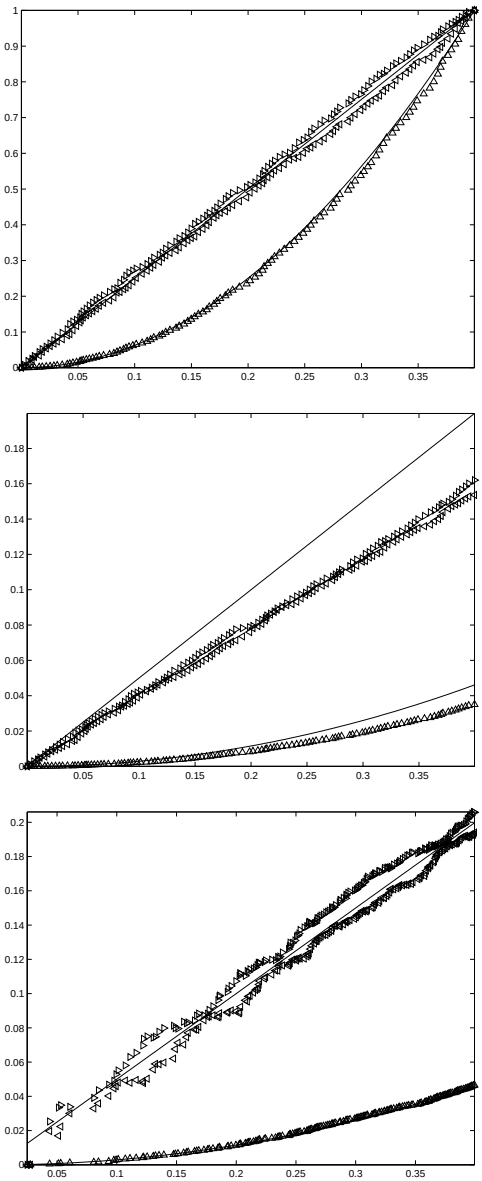

Fig. 4.10 Local SVD Curves for the unit sphere. Numerical values are triangles, analytic estimates are indicated by the solid line. Top: the spherical cap has a uniform distribution; Middle: the distribution of data was not uniform; Bottom: whitened computations, with the analytic estimates.

we present a method for eliminating the dependence of the singular values on the local data statistics [34]. The result will be that the singular values scale as a function of the dimension only. This is achieved by introducing a local whitening transform which leads to a normalized scaling of the singular value curves. The purpose of this scaling is to facilitate the determination of the nature of the scaling of the singular values. This is especially useful for applications which require that the dimension be estimated in a large number of local regions. For such problems it is desirable for the procedure to be as automatic as possible.

The basis for the modified algorithm is the observation that clusters defined by *spherical* balls do not reflect the distribution of the data. The information computed by the basic algorithm, however, provides a description of the local ball which includes the data distribution. Before developing the specifics of the algorithm which exploits this observation, we state and prove the following geometrical fact:

Proposition 4.4. *Let C be a nonsingular ensemble averaged covariance matrix generated by the $N \times P$ data matrix X, i.e., $C = XX^T$. Consider the ellipsoid defined by*

$$\mathbf{x}^T C^{-1} \mathbf{x} = 1 \tag{4.50}$$

The eigenvectors of C are the directions of the principal axes of the ellipse and the associated square-roots of the eigenvalues are the lengths of the semi-axes.

Proof. It suffices to trasform the equation for the ellipse into the KL basis. Let U consist of the eigenvectors of C. Then

$$\mathbf{z} = U^T \mathbf{x}$$

is the appropriate change of basis. Since U is orthogonal $\mathbf{x} = U\mathbf{z}$. Substituting this relation into Equation (4.50) gives

$$\mathbf{z}^T U^T C^{-1} U \mathbf{z} = 1 \tag{4.51}$$

Given that U diagonalizes C we have

$$C = U \Lambda U^T$$

Assuming C is nonsingular,

$$C^{-1} = U \Lambda^{-1} U^T$$

Substituting this equation into Equation (4.51) gives

$$\mathbf{z}^T U^T U \Lambda^{-1} U^T U \mathbf{z} = 1$$

Again, using the orthogonality of U, we have $\mathbf{z}^T \Lambda^{-1} \mathbf{z} = 1$ or

$$\sum_{i=1}^{N} \frac{z_i^2}{\lambda_i} = 1$$

Note that the expression $\mathbf{x}^T C^{-1}\mathbf{x} = c$ represents a family of concentric ellipses.

A similar proof may be found in [39]. See also [33]. □

Equipped with this geometrical fact, we return to the develop of the modified algorithm. The starting point is the same as for algorithm 4.7.4, i.e., the computation of the KL eigenvectors (left-singular vectors) and eigenvalues (singular values squared) as the radius of a local neighborhood changes. This information describes the distribution of the data in terms of a *hyperellipsoid.* Thus the local distribution of the data may be accounted for if ellipsoidal balls are employed in the scaling routine rather than spherical balls which treat the data as if it were uniformly distributed.

The ellipsoidal neighborhoods determined in the first stage of the algorithm may be used to order the data in the ϵ-neighborhood matrix via the introduction of the $A - norm$ written $\|\mathbf{x}\|_A^2 = \mathbf{x}^T A\mathbf{x}$ where the matrix A is taken to be C^{-1} and $C = B_\epsilon B_\epsilon^T$.

$$\begin{aligned} \|\mathbf{x}\|_{C^{-1}}^2 &= \mathbf{x}^T C^{-1}\mathbf{x} \\ &= \mathbf{x}^T U\Lambda^{-1}U^T\mathbf{x} \\ &= (\mathbf{x}^T U\Lambda^{-1/2})(\Lambda^{-1/2}U^T\mathbf{x}) \end{aligned}$$

Defining

$$\mathbf{y} = \Lambda^{-1/2}U^T\mathbf{x}$$

we see that we may write

$$\|\mathbf{x}\|_{C^{-1}}^2 = \mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|^2$$

Hence the ellipsoidal norm may be interpretted as the standard Euclidean metric after the change of coordinates

$$Y = \Lambda^{-1/2}U^T B_\epsilon$$

This transformation may be viewed as a *local whitening* transformation given that the covariance matrix is the identity in the new coordinate system, i.e.,

$$YY^T = I \tag{4.52}$$

We now observe that this local whitening transformation removes the dependence of the scaling on the statistics of the data, and creates singular values which depend exlusively on the local geometry of the data surface. In particular, all the singular values are now normalized to have value $\sigma_i = 1$ for the neighborhood of maximum radius $B_{\epsilon_{\max}}$.

Example 4.12. Local KL on the Sphere (revisited). In this example we applied the local whitening transformation to data nonuniformly sampled from the sphere (see the middle graph of Figure 4.10 for the results using the standard local KL algorithm). The bottom graph of Figure 4.10 shows the normalized singular values computed using the whitened data.

The primary result that we exploit is that the whitened data produces normalized singular value curves.

> *Geometric Scaling Law [34].* The slopes of all linear scaled singular values must be $\frac{1}{\epsilon_{\max}}$, and the quadratic coefficient must be $\frac{1}{\epsilon_{\max}^2}$.

In other words, for eigenvectors which span the tangent space, their associated local SVD curves must have the form

$$\sigma_j(\epsilon) = \frac{\epsilon}{\epsilon_{\max}}$$

Similarly, the region of lowest order curvature of the function has local SVD curves of the form

$$\sigma_j(\epsilon) = \frac{\epsilon^2}{\epsilon_{\max}^2}$$

By using the hyperellipsoidal metric, or equivalently, the Euclidean metric on locally whitened data, we have constructed a standardization of the slopes of the local singular value curves. This standardization has an important algorithmic consequence which we now discuss.

4.7.5.1 Efficient Implementation of the Geometric Scaling Law. The local dimensionality estimation algorithm based on the energy criterion, as well as the algorithm based on the scaling criterion require that many SVD calculations be carried out for each local region. Following [34], we now describe how the transformation described above permits the local dimension to be estimated in a fashion which requires only one SVD calculation per local region.

The idea is that the normalized line with $y_1(x) = x/\epsilon_{\max}$ and a quadratic $y_2(x) = x^2/\epsilon_{\max}^2$ have greatest deviation at the midpoint $x = \epsilon_{\max}/2$. Given $y_1(\epsilon_{\max}/2) = 1/2$ and $y_2(\epsilon_{\max}/2) = 1/4$ it follows that if $\sigma_i > 3/8$ the line model is more likely while if $\sigma_i < 3/8$ the quadratic (or higher order) model is more likely. Thus, in determining which dimensions are scaling linearly, we need make only s single computation, at a distance of $\epsilon_{\max}/2$ from center, where the distinction between linear and higher order scaling is the greatest. The decision is thus:

> *Geometric Scaling Dimensionality Criterion:* If $\sigma_i(\epsilon_{\max}/2) > \frac{3}{8}$, then eigenvector i belongs to the span of the tangent plane. The local dimension is the number singular values for which this condition is satisfied.

These considerations may be translated into an efficient algorithm which requires, ideally, only one eigenvector computation.

Fast Normalized Scaling Criterion Algorithm

- Compute the local KL eigenvectors and singular values of $B_{\epsilon_{\max}}$.
- Locally whiten the data.
- Compute the singular values of $B_{\epsilon_{\max}/2}$.
- Determine the number of singular values such that $\sigma_i(\frac{\epsilon_{\max}}{2}) > 3/8$.

```
This is the local dimension.
```

The intialization of $\epsilon_{\max}$ may be done by requiring that a maximum number of data points be included in the neigborhood. In practice, the minimum value of ϵ may also be specified by requiring a minimum number of points in B_ϵ.

Nonlinear Generalizations

While the global Karhunen-Loève expansion is an extremely powerful tool, it is limited by the fact that it is a linear procedure. The local KL algorithm described above is one approach to overcoming the limitations of such linear methods. In Chapter 9 several partial or fully nonlinear reduction architectures will be proposed.

Problems

Questions 4.1 - 4.9 deal particularly with ensembles extended according to Definition 4.1.

4.1 Show that the function $(f(x)+f(-x))/2$ is even and the function $(f(x)-f(-x))/2$ is odd.

4.2 Show that the ensemble average $\langle u(x) \rangle$ of a data set which is extended is even.

4.3 Show that the symmetrized kernel is even in both variables, i.e.,

$$\hat{C}(x,y) = \hat{C}(-x,-y).$$

4.4 $C_e(x,y)$ is even in x and y and $C_o(x,y)$ is odd in x and y.

4.5 Show that the solutions of

$$\int C_e(x,y)\psi(y)dy = \lambda\psi(x)$$

must be even functions and that the solutions of

$$\int C_o(x,y)\psi(y)dy = \lambda\psi(x)$$

must be odd functions.

4.6 The even and odd eigenfunctions are orthogonal, i.e., for any ϕ_e and ϕ_o we have

$$(\phi_e, \phi_o) = 0.$$

4.7 Show that all the even eigenfunctions $\hat{\phi}_e$ belong to the null-space of $\hat{C}_o$ and that all the odd eigenfuctions $\hat{\phi}_o$ belong to the null-space of $\hat{C}_e$.

4.8 Derive Equation (4.17).

4.9 Consider again the data matrix

$$X = \begin{pmatrix} -2 & -1 & 1 \\ 0 & -1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix}$$

The matrix of data which corresponds to the *reflection* of the data matrix above is then

$$X_R = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 2 \\ 0 & -1 & 0 \\ -2 & -1 & 1 \end{pmatrix}$$

Hence the symmetry extended ensemble consists of the data matrix

$$\tilde{X} = \begin{pmatrix} -2 & -1 & 1 & 1 & -1 & 1 \\ 0 & -1 & 0 & -1 & 1 & 2 \\ -1 & 1 & 2 & 0 & -1 & 0 \\ 1 & -1 & 1 & -2 & -1 & 1 \end{pmatrix}$$

Show that the data matrix X may be written as the sum of an even 4×3 matrix X_e and an odd 4×3 matrix X_o.

a) Compute the eigenvectors and eigenvalues of $\tilde{X}\tilde{X}^T$, i.e., the best basis for the columns of $\tilde{X}$.

b) Compute the eigenvectors and eigenvalues of $X_e X_e^T$ and $X_o X_o^T$ and compare with those computed in part a) and problem 3.21.

c) Comment on the form of the eigenvectors in the nullspaces of the eigenvector problems solved in part b). Keep in mind that eigenvectors corresponding to a multiple eigenvalue are not uniquely defined.

d) Decompose the eigenvector of XX^T corresponding to the largest eigenvalue into even and odd components $\mathbf{u} = \mathbf{u}_e + \mathbf{u}_o$. Under what conditions are $\mathbf{u}_e, \mathbf{u}_o$ eigenvectors?

4.10 Show that if $\phi(x)$ is an eigenfunction of equation (4.8) with eigenvalue λ then $\pm R\phi(x)$ is also an eigenfunction of equation (4.8) with eigenvalue λ.

4.11 Prove the orthogonality relations given by Equation (4.24).

4.12 Consider the two eigenvector problems

$$C_{\mathbf{x}}\mathbf{u} = \lambda_x \mathbf{u}$$

and

$$C_{\mathbf{s}}\mathbf{v} = \lambda_s \mathbf{v}$$

where the matrices are related by $C_{\mathbf{x}} = C_{\mathbf{s}} + \alpha I$ where α is a real number and I is the identity matrix. Show that if $\mathbf{u}$ is and eigenvector of $C_{\mathbf{x}}$, then it is also an eigenvector of $C_{\mathbf{s}}$ with eigenvalue $\lambda_s = \lambda_x - \alpha$.

4.13 Let A be a real $m \times n$ matrix. Show that the matrix M defined as

$$M = \alpha^2 I + AA^T$$

is non-singular where I is the $m \times m$ identity matrix.

4.14 Compute the KL eigenvectors and eigenvalues of the graphs of the functions

1. $y = x^2$
2. $y = x^4$

about the point $(x, y) = (1, 1)$, i.e., on the domain interval $(1 - \epsilon, 1 + \epsilon)$. In each case,

- Identify how the eigenvalues and eigenvectors depend ϵ.
- Determine the Taylor expansions and compare the range of the Jacobian matrix with the results found above for the KL eigenvectors.

4.15 Compute the KL eigenvectors and eigenvalues for graphs of the family of curves

$$y = \kappa x^n$$

on the interval $(-\epsilon, \epsilon)$. Compare your results with the Taylor series expasion about $x = 0$.

4.16 This problem concerns completing the calculations summarized in Exercise 4.11. In particular, it requires you to compute a local basis for unit sphere centered at the origin. We choose arbitrarily to compute the basis about the north pole $(0, 0, 1)$.

$$T_\epsilon = [0, 2\pi] \times [\pi, \pi + \epsilon]$$

- Given the formula for the surface area S of the patch

$$S = \int_{T_\epsilon} \int \| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \| du dv$$

show that

$$S = 2\pi \cos(\epsilon - 1)$$

- Given

$$\int_S \int f dS = \int_{T_\epsilon} \int f(\mathbf{x}(u, v)) \| \frac{\partial \mathbf{x}}{\partial u} \times \frac{\partial \mathbf{x}}{\partial v} \| du dv$$

show that

$$C_{ij} = \frac{1}{S}\int_{\pi}^{\pi+\epsilon}\int_{0}^{\pi} x_i(u,v)x_j(u,v)|\cos v du dv$$

- Show that
  - $C_{11} = C_{22} = \frac{1}{3} - \frac{1}{6}\cos\epsilon(1+\cos\epsilon)$
  - $C_{33} = \frac{1}{3} - \frac{2}{3}\cos\epsilon - \frac{1}{6}\cos^2\epsilon)$
  - $C_{ij} = 0$ if $i \neq j$
- From the Law of Cosines we have $R = 1 - \cos\epsilon$ where R is the radial distance from the point $(0,0,1)$. Show that the eigenvalues may be rewritten as a function of R to give $\lambda_1 = \lambda_2 = R^2/4 - R^4/24$ and $\lambda_3 = R^4/12$.

4.17 Consider the vector function

$$\mathbf{f}(x,y) = \begin{pmatrix} x \\ y \\ x^2+y^2 \end{pmatrix}$$

in the square local region $x, y \in (1-\epsilon, 1+\epsilon)$.

(a) Compute the linear approximation $d\mathbf{f}$ about the point $\mathbf{x}_o = (1,1)^T$. Are the columns of the Jacobian orthogonal?

(b) Compute the local KL eigenvectors and eigenvalues centered about the point $\mathbf{f} = (1,1,2)^T$ and compare with part (a). To solve this problem, you may either use a symbolic math package or procede numerically. In the latter case, evaluate the ensemble average covariance matrix as a function of ϵ and compute the eigenvalues and eigenvectors for several different values of ϵ. Plot the square roots of the eigenvalues versus ϵ. Determine the eigenvectors which span the tangent space, i.e., the range of the Jacobian, based on the observed scaling.

(c) Using techniques from Calculus, compute the plane tangent to the surface $z = f(x,y) = x^2 + y^2$ at the point $(x,y) = (1,1)$. Show that this is the same plane as the one spanned by the bases in parts (a) and (b).

(Hint: Compare with Problem 4.16. Also, note the following: Given a vector in the plane $\mathbf{x}_1$ and the vector normal to the plane $\mathbf{n}$, the equation for the plane is given by $(\mathbf{x} - \mathbf{x}_1)\cdot\mathbf{n} = 0$. Two vectors $\mathbf{x}_1, \mathbf{x}_2$ span a plane which has as its normal the vector cross product $\mathbf{n} = \mathbf{x}_1 \times \mathbf{x}_2$. Consider the surface $z = f(x,y)$. The plane through the point $\mathbf{r}_o$ having $\mathbf{n}$ as its normal is called the tangent plane to the surface at $\mathbf{r}_o$. The normal vector is given by $\mathbf{n} = -f_x\mathbf{i} - f_y\mathbf{j} + \mathbf{k}$.)

4.18 Prove Equation (4.52).

4.19 Proposition 4.4 assumes that the covariance matrix C is non-singular. Restate the proposition for the case that C is singular. Adapt the proof in the notes for this case.

Computer Projects

4.20 The project concerns the application of the KL procedure for incomplete data [19]. Let the complete data set be translationally invariant

$$f(x_m, t_\mu) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} \sin(k(x_m - t_\mu)) \tag{4.53}$$

where $m = 1, \ldots, M$ is the dimension of the ambient space (size of the spatial grid) and $\mu = 1, \ldots, P$ is the number of points in the ensemble. Let $x_m = (m-1)2\pi/M$ and $t_\mu = (\mu - 1)2\pi/P$. Select an ensemble of masks $\{\mathbf{m}^{(\mu)}\}$, $\mu = 1, \ldots, P$, where 10% of the indices are selected to be zero for each mask. Each pattern in the incomplete ensemble may be written

$$\tilde{\mathbf{x}}^{(\mu)} = \mathbf{m}^{(\mu)}.\mathbf{f}^{(\mu)}$$

where $(\mathbf{f}^{(\mu)})_m = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} \sin(k(x_m - t_\mu))$. Let $P = M = 64$ and $N = 3$.

1. Compute the eigenvectors of this ensemble using the algorithm 4.5.2.
2. Plot the eigenvalues as a function of the iteration and continue until they converge.
3. Plot your final eigenfunctions corresponding to the 10 largest eigenvalues.
4. Plot the element $\tilde{\mathbf{x}}^{(1)}$ and the vector $\tilde{\mathbf{x}}_D$ repaired according to Equation (4.42). Determine the value of D which gives provides the best approximation to the original non-gappy pattern vector.

Problem 4.1. *This project concerns the computer implementation of local KL.*

Part III

Time, Frequency and Scale Analysis

5

Fourier Analysis