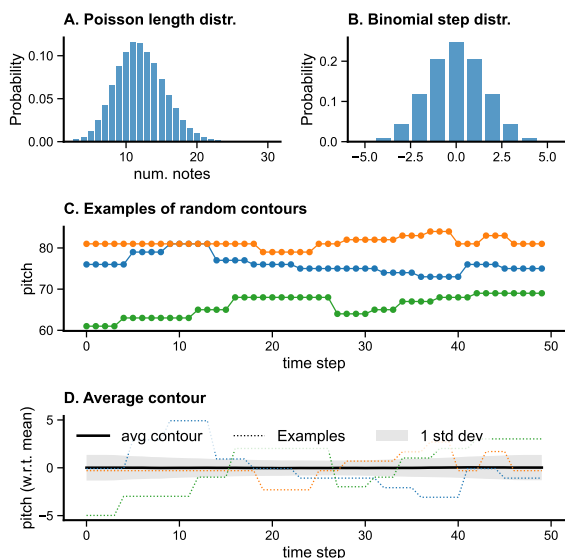


Cosine Contours
A multipurpose representation
for melodies

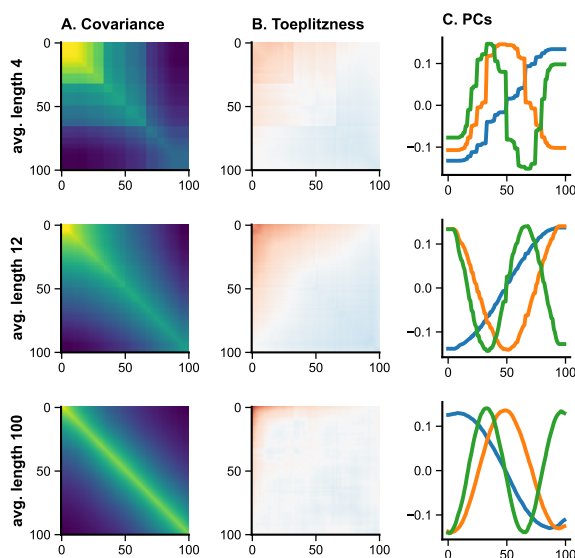
SUPPLEMENTARY MATERIALS

Bas Cornelissen, Willem Zuidema & John Ashley Burgoyne

S1 Random walk baseline



We compared the principal components of phrases to a random walk baseline that was intended to be fairly similar to actual phrase contours. First, we draw the length (number of notes) K of the random walk from a Poisson distribution with mean $\lambda = 12$ (truncated below 3). The value 12 was chosen so as to approximate the length distribution of phrases. Then we draw an initial pitch x_0 uniformly between 60 and 85 (in MIDI pitch space). Next, at every step k we draw the size of a step r_k (the interval) from a Binomial distribution with parameters $n = 10$ and $p = 0.5$, shifted to have mean 0, and let the next pitch be $x_k = x_{k-1} + r_k$. We constrain the step sizes to lie between -12 and $+12$, meaning that jumps cannot exceed an octave. This results in small, approximately normally distributed step sizes. This process yields a sequence of pitches x_0, \dots, x_{K-1} . As usual, we interpolate a step function through these pitches and sample $N = 100$ equally spaced pitches to obtain a random contour. In the figure above we use $N = 50$ for readability.



Here we vary the average length λ of the random walk baseline. This affects the number of notes K , but we still have $N = 100$ throughout. We generate 10,000 random contours, and compute the covariance matrices (A). The longer the melodies (larger K), the more the covariance matrix starts to resemble a Toeplitz matrix, which has constant values along each of its diagonals. As an ad-hoc measure of *Toeplitzness*, we measure how much every entry of the covariance matrix differs from the mean value on that diagonal. For a Toeplitz matrix, that should be zero everywhere: all diagonals are constant, so every entry also equals the mean of that diagonal. Column (B) makes clear that the covariance matrix differs from a Toeplitz matrix mostly in the upper left corner, which contains the covariance in the first timesteps. All this is also reflected in the principal components (C).

S2 Analyses of other datasets

In this section we visualize the principal components of melodic material from motifs to songs in different traditions. For every dataset we show:

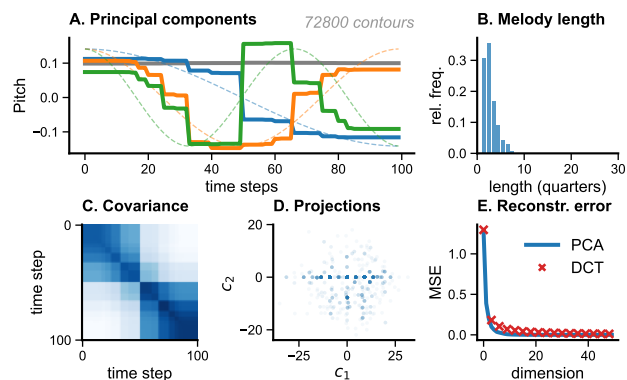
- The first four principal components. The first one is usually a flat line (gray), the second a descending shape (blue), the third a convex shape (orange), and the fourth one undulating (green). The corresponding cosines are shown as thin dashed lines in the same colors.
- The length distribution of the melodic material, where length is measured in quarter notes. For Gregorian chant we assume all notes are quarter notes.
- The covariance matrix.
- A scatterplot showing the representations of 2000 contours in 2d cosine contour space.
- The reconstruction error using the discrete cosine transform compared to a principal component analysis.

It is clear that the cosine approximation is most accurate at the phrase level. For very short melodic fragments (such as neumes or syllables), you see clear effects of the typical number of notes. For example, neumes often have only 2 notes, meaning there is a jump in the middle of the contour. You can see this in the principal components, but also in the covariance matrix. Such effects are weaker, but sometimes still visible at the phrase level: German folksongs apparently often have durations of 8 quarter notes, with jumps in the middle, or after 2 of 6 quarter notes. For complete songs, finally, the principal components are often difficult to interpret. Only for a very large number of songs (such as when combining all chants in GregoBase) does a pattern reminiscent of the cosines emerge. But for very small datasets, such as those in the Densmore collection, the principal components are very irregular.

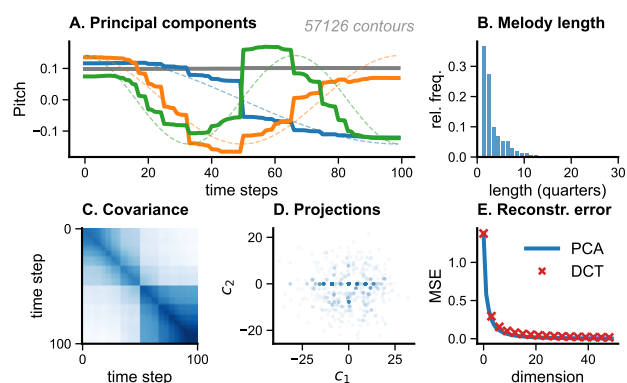
S2.1 Motifs

All motifs come from Gregorian chant (responsories from CantusCorpus).

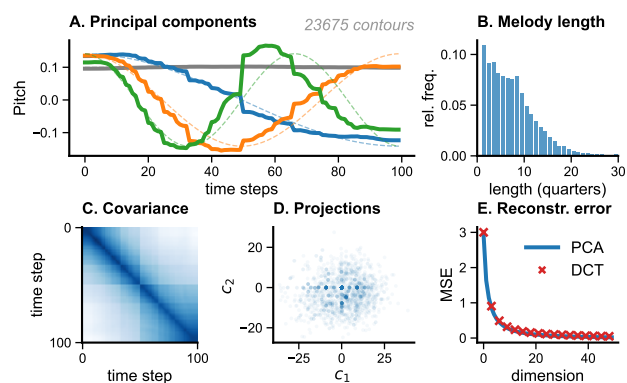
Neumes



Syllables

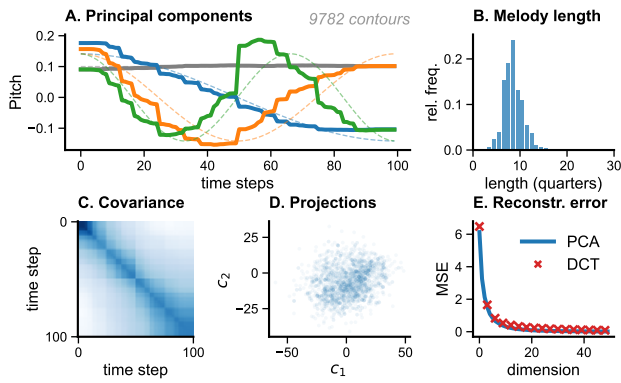


Words



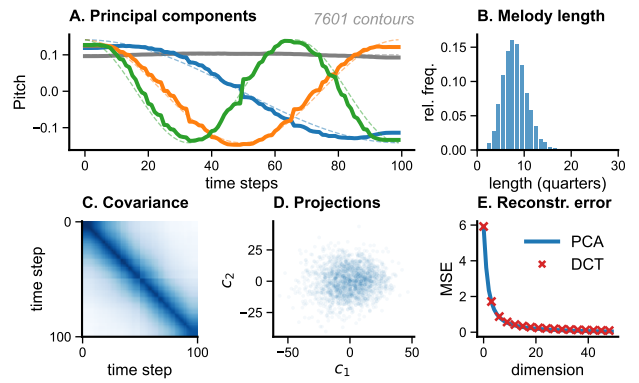
S2.2 Phrases

German: Erk

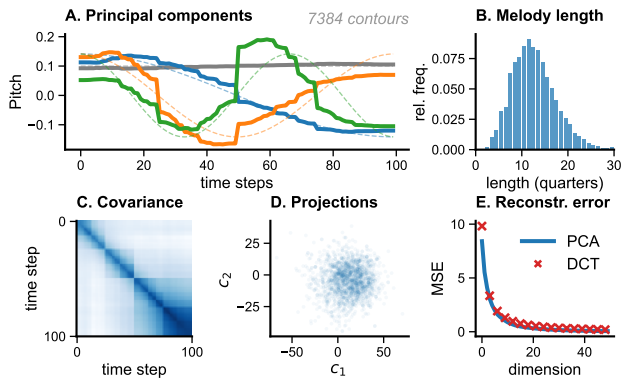


S2.3 Random segments

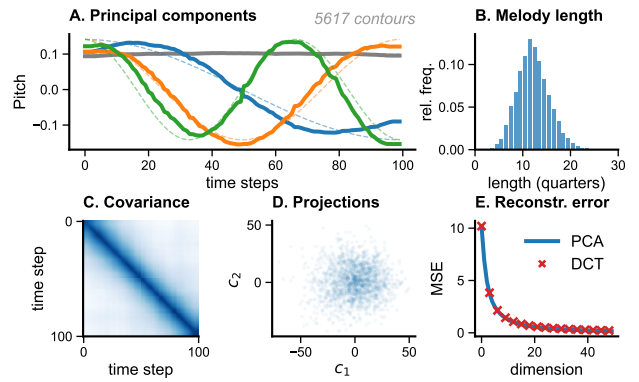
German: Erk



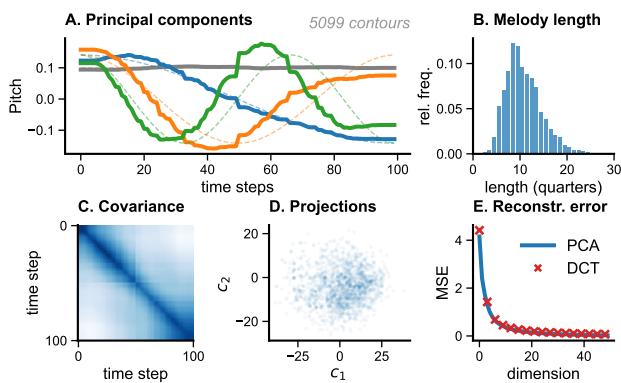
Chinese: Han



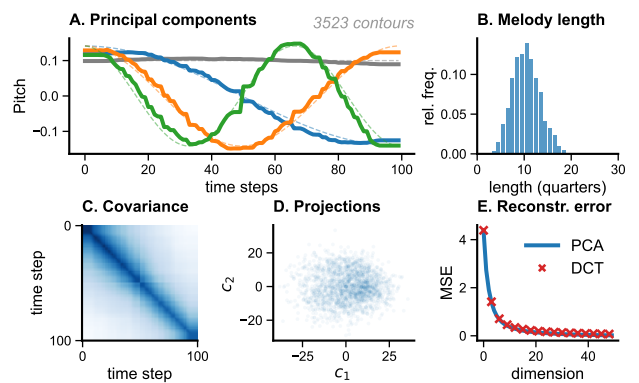
German: Han



Antiphons

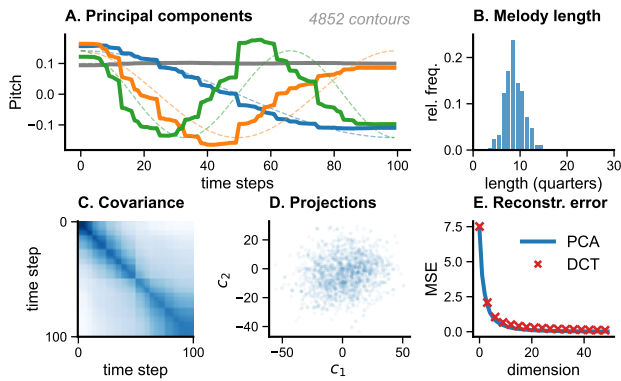


Antiphons



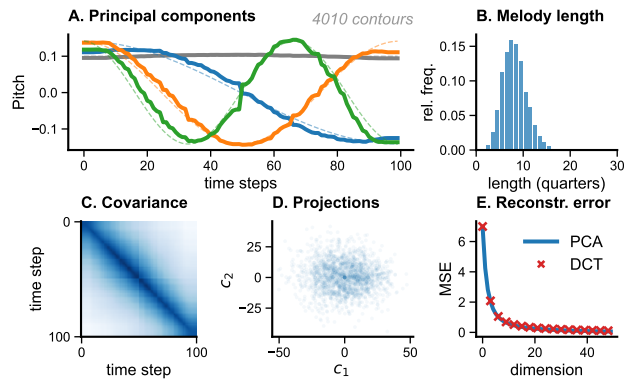
S2.4 Phrases (continued)

German: Boehme

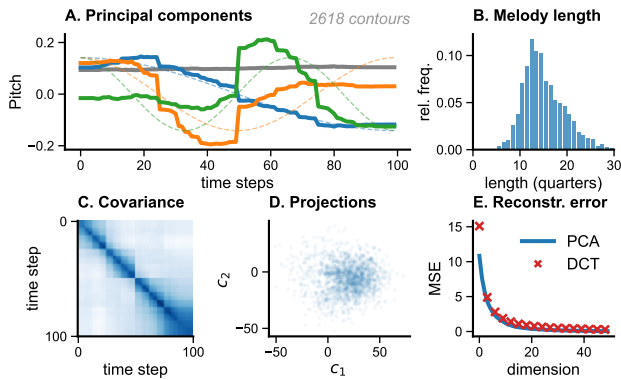


S2.5 Random segments (continued)

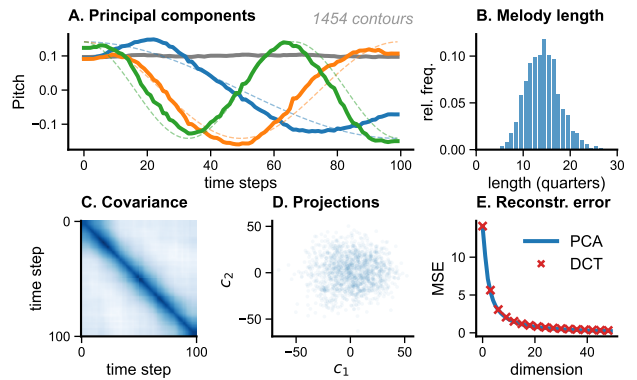
German: Boehme



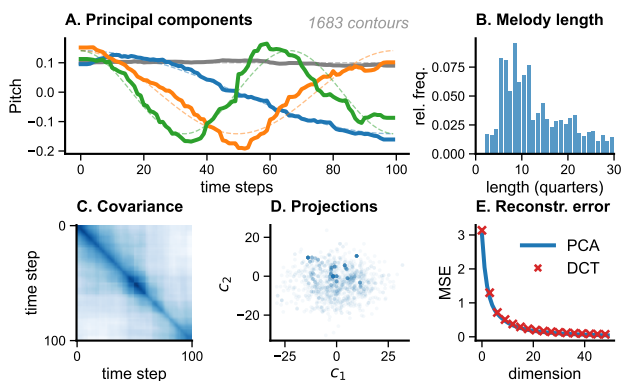
Chinese: Shanxi



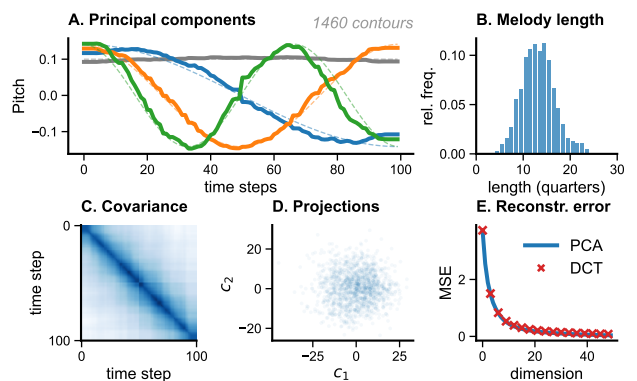
Chinese: Shanxi



Responsories

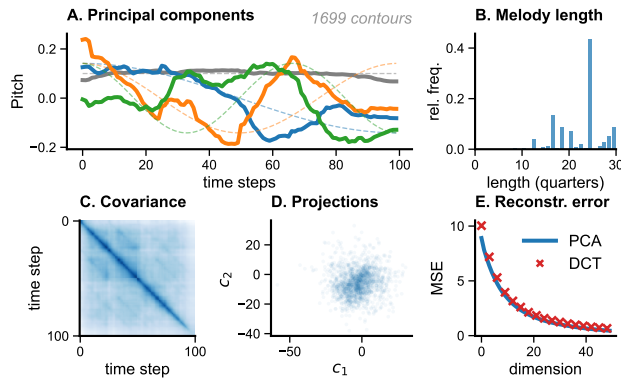


Responsories

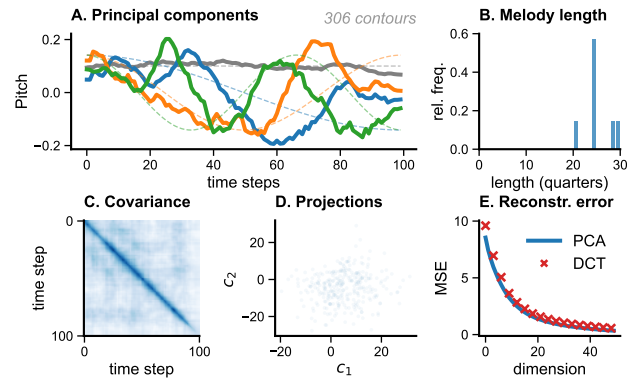


S2.6 Songs

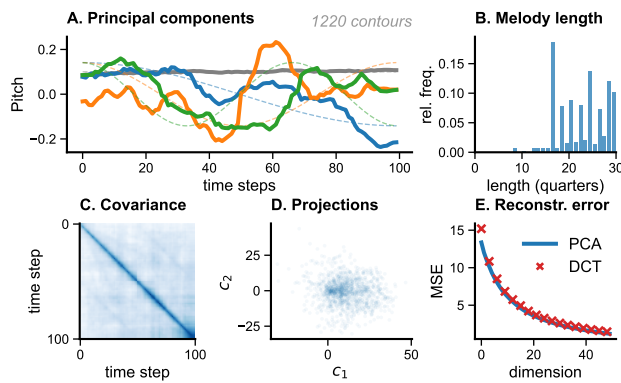
German: Erk



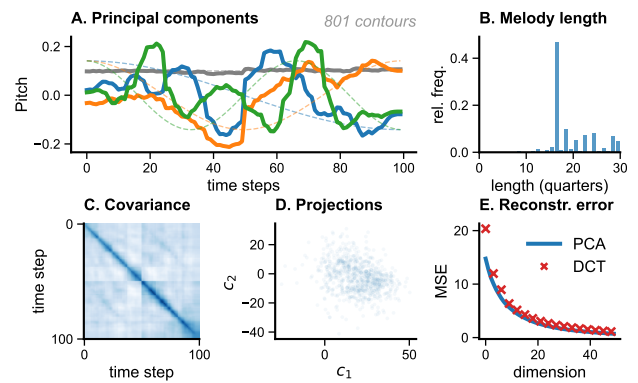
German: Boehme



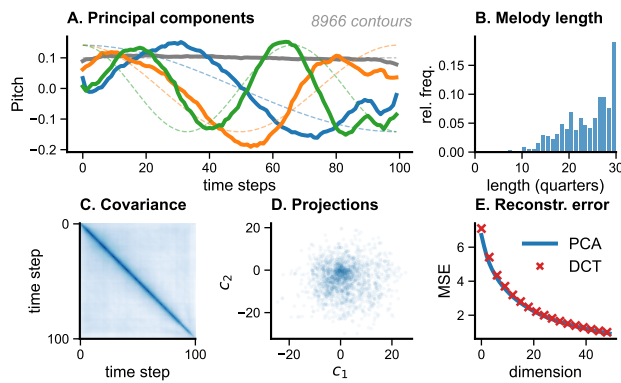
Chinese: Han



Chinese: Shanxi



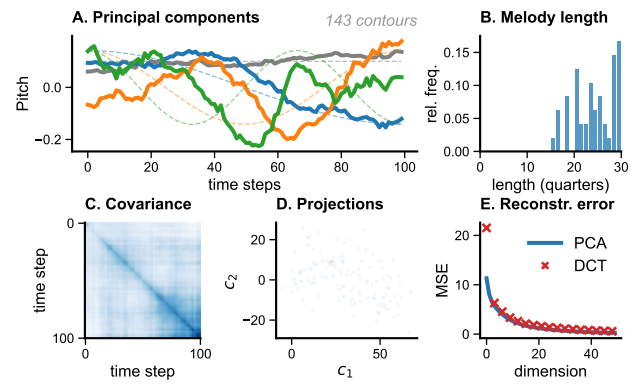
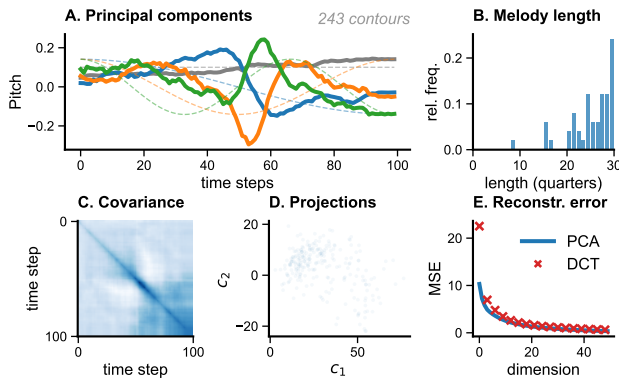
All chants in GregoBase



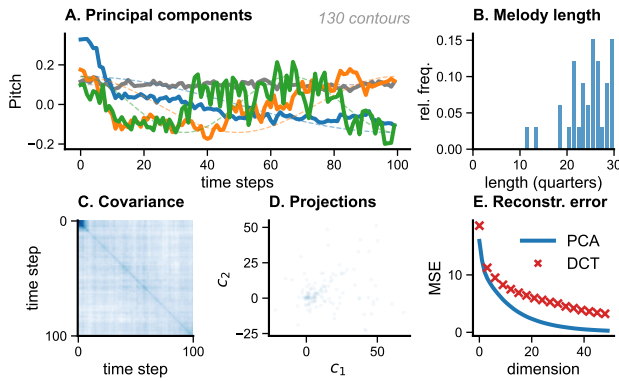
S2.7 Songs (continued)

Menominee (Densmore)

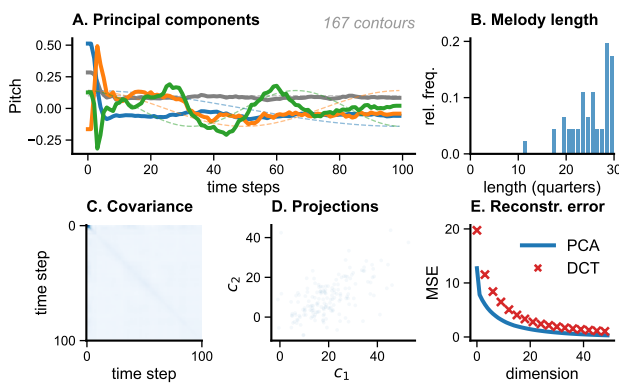
Teton Sioux (Densmore)



Nootka (Densmore)



Papago (Densmore)



S3 Mathematical background

In this section we provide some more mathematical background to illustrate why we observe cosine-shaped principal components. The aim is to make some of the key points a bit more accessible; we refer to Jolliffe (2002) for a detailed discussion of principal component analysis, to Gray (2006) for a rigorous treatment of Toeplitz matrices and their limiting behaviour, and to Rao and Yip (1990) for the discrete cosine transform.

Notation We write N for the length of a contour, or the number of steps in a random walk, and M denotes the number of contours. Consider a dataset $\{\mathbf{x}_1, \dots, \mathbf{x}_M\}$ of points $\mathbf{x}_m = (x_{m1}, \dots, x_{mN})$ in \mathbb{R}^N . We denote the sample mean by $\bar{\mathbf{x}}$ and the centered data points by $\hat{\mathbf{x}}_m$:

$$\bar{\mathbf{x}} = \frac{1}{M} \sum_{m=1}^M \mathbf{x}_m \quad \text{and} \quad \hat{\mathbf{x}}_m = \mathbf{x}_m - \bar{\mathbf{x}}, \quad (\text{S3.1})$$

and both of course live in \mathbb{R}^N . An $M \times N$ matrix \mathbf{X} has entries $(\mathbf{X})_{m,n} = x_{mn}$, and for $N \times N$ matrices we generally index rows by n and columns by k .

S3.1 Principal components

Maximize projected variance The goal of a principal component analysis is to find a subspace of lower dimensionality $D < N$ that maximizes the variance of the data when it is projected on this subspace. First, we project the data on a one-dimensional subspace spanned by the unit vector $\mathbf{u}_1 \in \mathbb{R}^N$. You can think of the projection of \mathbf{x}_n as a point in the N -dimensional ambient space, but we rather treat it as the scalar $\mathbf{u}_1^T \mathbf{x}_n$: the coordinate in the one-dimensional subspace. The projected data then has mean $\mathbf{u}_1^T \bar{\mathbf{x}}$ and variance

$$\frac{1}{M} \sum_{m=1}^M (\mathbf{u}_1^T \mathbf{x}_m - \mathbf{u}_1^T \bar{\mathbf{x}})^2 = \mathbf{u}_1^T \mathbf{S} \mathbf{u}_1, \quad (\text{S3.2})$$

where \mathbf{S} is the $N \times n$ covariance matrix given by

$$\mathbf{S} = \frac{1}{M} \sum_{m=1}^M (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T \quad (\text{S3.3})$$

We want to choose \mathbf{u}_1 in such a way that it maximizes the projected variance $\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1$. It can be shown, using a Lagrange multiplier, that under the constraint $\|\mathbf{u}_1\| = 1$, the projected variance is maximized when

$$\mathbf{S} \mathbf{u}_1 = \lambda_1 \mathbf{u}_1 \quad (\text{S3.4})$$

(Jolliffe, 2002, see e.g.). Left-multiplying by \mathbf{u}_1^T , and using that $\mathbf{u}_1^T \mathbf{u}_1 = 1$, this is the case when

$$\mathbf{u}_1^T \mathbf{S} \mathbf{u}_1 = \lambda_1. \quad (\text{S3.5})$$

Equation (S3.4) shows that \mathbf{u}_1 must be an eigenvector of the covariance matrix \mathbf{S} corresponding to eigenvalue λ_1 , which is exactly the projected variance according to (S3.5). The first principal component, in short, is the eigenvector of the covariance matrix corresponding to the largest eigenvalue. The argument can be extended inductively to identify all principal components as eigenvectors of the covariance matrix, ordered according to their eigenvalues.

Minimize reconstruction error It should be noted that one can also motivate principal components in another way. Consider a dataset $\{\mathbf{x}_m \in \mathbb{R}^N\}_m$ as before, and a set of basis vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_D\}$ for \mathbb{R}^N with norm 1. As before, the projection of \mathbf{x} on the \mathbf{u}_n is $c_n = \mathbf{u}_n^T \mathbf{x}$, and so we can represent \mathbf{x} as a coordinate vector (c_1, \dots, c_D) . Now suppose we only use the first D coordinates to represent \mathbf{x} , so we get the truncated representation:

$$\bar{\mathbf{x}} = \sum_{i=1}^D c_i \mathbf{u}_i. \quad (\text{S3.6})$$

Now measure the *reconstruction error* as

$$\text{MSE} = \frac{1}{M} \sum_{m=1}^M (\mathbf{x}_m - \bar{\mathbf{x}})^2 \quad (\text{S3.7})$$

We ask: how should we choose the basis vectors so that the reconstruction error MSE is minimized? The answer is the same: as the eigenvectors, ranked in descending order (Rao & Yip, 1990).

S3.2 Toeplitz and circulant matrices

Toeplitz matrices are matrices where every diagonal has the same value. They are usually indexed as follows:

$$\mathbf{T} = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \dots & t_{-(N-1)} \\ t_1 & t_0 & t_{-1} & & \\ t_2 & t_1 & t_0 & & \\ \vdots & & & \ddots & \\ t_{N-1} & & & \dots & t_0 \end{bmatrix} \quad (\text{S3.8})$$

That means that $T_{i,j} = t_{j-i}$. Before we discuss Toeplitz matrices further, let's focus on the special subset of circulant matrices. A *circulant matrix* is a Toeplitz matrix

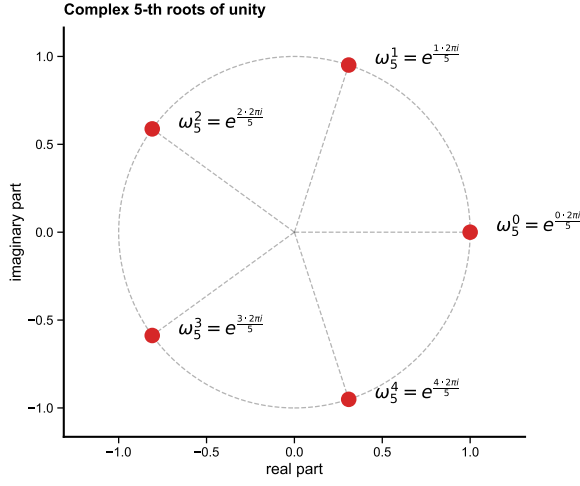


Figure S3.1: The N -th roots of unity for $N = 5$ are points on the complex unit circle.

where every row equals the previous row, rotated one step to the right:

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{N-1} \\ c_{N-1} & c_0 & c_1 & & c_{N-2} \\ c_{N-2} & c_{N-1} & c_0 & & \\ \vdots & & \ddots & & \vdots \\ c_1 & & & c_0 & c_1 \\ & & & \dots & c_{N-1} & c_0 \end{bmatrix} \quad (\text{S3.9})$$

It is convenient to start indexing at 0 rather than 1, so that we have $\mathbf{C}_{n,k} = c_{k-n \bmod N}$. We read the subscripts periodically, so that e.g. $c_{N+3} = c_3$. For circulant matrices, matrix multiplication takes the form of a *circular convolution*: if $\mathbf{y} = \mathbf{C}\mathbf{x}$, we have

$$y_n = \sum_{k=0}^{N-1} c_{k-n} x_k. \quad (\text{S3.10})$$

Eigenvectors of circulant matrices Surprisingly, all circulant matrices have the same eigenvectors. These eigenvectors consist of (N -th) roots of unity: the complex numbers z satisfying $z^N = 1$. The first complex root of unity is

$$\omega = e^{\frac{2\pi i}{N}}, \quad (\text{S3.11})$$

and its powers ω^k are other roots of unity, since $(\omega^k)^N = (\omega^N)^k = 1$. The numbers $\omega^0, \dots, \omega^{N-1}$ can be visualized as evenly spaced points on the unit circle in the complex plane (see figure S3.1). Importantly, these numbers (like the coefficients c_k) are periodical: $\omega^{N+k} = \omega^N \cdot \omega^k = \omega^k$.

This property allows us to show that the N eigenvectors of a circulant matrix are

$$\boldsymbol{\omega}_n = (\omega^{n \cdot 0}, \dots, \omega^{n \cdot (N-1)})_{,,} \quad (\text{S3.12})$$

for $n = 0, \dots, N-1$. You can verify this directly when $n = 0$, since $\boldsymbol{\omega}_0$ is then an all-ones vector, but let's consider the general case. We have to show that $\mathbf{C}\boldsymbol{\omega}_n = \lambda_n \boldsymbol{\omega}_n$ for some constant λ_n . Using (S3.10), we can show that k 'th entry of the left hand side indeed equals $\lambda_n \omega^{nk}$:

$$(\mathbf{C}\boldsymbol{\omega}_n)_k = \sum_{j=0}^{N-1} c_{j-k} \cdot \omega^{n \cdot j} \quad (\text{S3.13})$$

$$= \omega^{nk} \cdot \sum_{j=0}^{N-1} c_{j-k} \cdot \omega^{n(j-k)} \quad (\text{S3.14})$$

$$= \omega^{nk} \cdot \underbrace{\sum_{j'=0}^{N-1} c_{j'} \cdot \omega^{n \cdot j'}}_{\lambda_n}. \quad (\text{S3.15})$$

Here we first multiplied by $\omega^{-nk}/\omega^{-nk}$ to align the indices of the coefficients and the powers. Then we used the periodicity of the roots of unity to reorder the sum, so it no longer depends on k and must equal the eigenvalue λ_n . The general case is similar.

Summarizing, every $N \times N$ circulant matrix \mathbf{C} has the same N eigenvectors $\boldsymbol{\omega}_0, \dots, \boldsymbol{\omega}_n$, with (different) corresponding eigenvalues:

$$\lambda_n = c_0 \omega^0 + c_1 \omega^n + \dots + c_{N-1} \omega^{n(N-1)} \quad (\text{S3.16})$$

$$= \sum_{j=0}^{N-1} c_j e^{\frac{2\pi i \cdot n j}{N}}, \quad (\text{S3.17})$$

for $n = 0, \dots, N-1$. From the second expression one sees that the eigenvalues $(\lambda_0, \dots, \lambda_{N-1})$ are the discrete Fourier transform of (c_0, \dots, c_{N-1}) .

Real circulant matrices In the scenario we are interested in, the matrix \mathbf{C} is real and symmetric, and such matrices have real eigenvalues and eigenvectors. To see that the eigenvalues are real, first note that a symmetric circulant matrix satisfies the additional constraint $c_k = c_{N-k}$. Also observe that ω^k and $\omega^{N-k} = \omega^{-k}$ are each others mirror image in the real axis (see figure S3.1). They have the same real part,

$$\text{Re}(\omega^k) = \cos\left(\frac{2\pi k}{N}\right), \quad (\text{S3.18})$$

and when adding them, the complex part cancels out: $\omega^k + \omega^{-k}$ lies on the real axis, at the point $2\text{Re}(\omega^k)$. This means that

$$c_k \omega^k + c_{N-k} \omega^{N-k} = 2c_k \text{Re}(\omega^k) \quad (\text{S3.19})$$

is a real number. From (S3.16) we see that the eigenvalues λ_n consist of many such sums: all complex parts cancel out and the eigenvalues are real¹

Now we can also choose real eigenvectors: the real part of $\boldsymbol{\omega}_n$. After all, if $\boldsymbol{\omega}_n$ is an eigenvector for the

¹The expression for the eigenvalues is slightly different depending on whether N is odd or even.

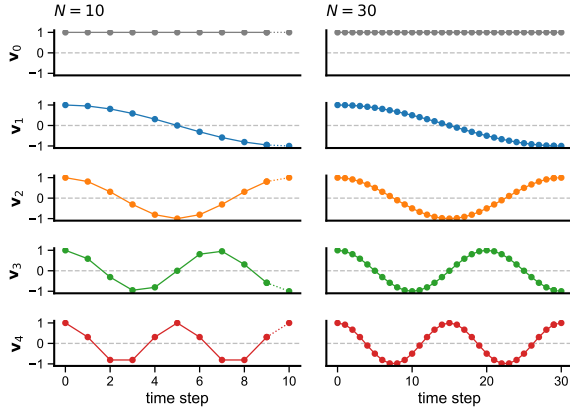


Figure S3.2: The eigenvectors of a symmetric, circulant matrix are discrete cosine functions with different periods.

real eigenvalue λ_n , so are ω_{-n} and $\mathbf{v}_n = 1/2(\omega_n + \omega_{-n})$. By the same argument as before, equations (S3.19) and (S3.18) show that this is a real eigenvector:

$$\mathbf{v}_n = \left(1, \cos \theta, \dots, \cos N\theta\right), \quad \theta = \frac{2\pi n}{N}. \quad (\text{S3.20})$$

This is a discrete cosine function consisting of N points, where higher n implies in higher frequencies. This is illustrated in figure S3.2.

Toeplitz is asymptotically circulant The reason circulant matrices are interesting here, is that Toeplitz matrices can be shown to be asymptotically equivalent to circulant matrices, and that eigenvalues are preserved. We refer to Gray (2006) for a detailed discussion of that result. What this implies is that the eigenvectors of large Toeplitz matrices are well approximated by those of circulant matrices: sinusoidal functions. That in turn means that approximately Toeplitz covariance matrices (which are real and symmetric) will have cosine-shaped eigenvectors.

S3.3 PCs of random processes

We want to end by discussing two examples where Toeplitz covariance structures arise, and we thus would expect cosine eigenvectors, at least asymptotically.

Weakly stationary process Toeplitz matrices arise naturally in the study of weakly stationary processes. These are random processes where the mean is constant over time, and where the covariance does not change by shifts in time: it only depends on the distance between two time steps. That is, when $\text{Cov}(x_i, x_j) = K(j - i)$ is some function of $j - i$, and thus results in a Toeplitz covariance matrix.

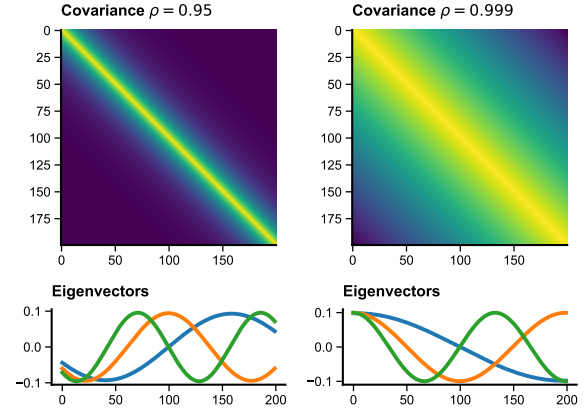


Figure S3.3: The autocovariance matrix for an autoregressive process AR(1) for two values of ρ . When $\rho \rightarrow 1$ it approximates the discrete cosine transform.

One example of such a process is a first order autoregressive process AR(1), where

$$x_n = \rho x_{n-1} + r_n, \quad (\text{S3.21})$$

where r_n is a random step with mean zero and variance σ^2 , and we assume $x_0 = 0$. It can be shown that this process has mean $E[x_n] = 0$ and variance $\text{Var}[x_t] = \sigma^2 / (1 - \rho^2)$ if $|\rho| < 1$. In that case, the covariance is

$$\text{Cov}(x_i, x_j) = \frac{\sigma^2}{1 - \rho^2} \cdot \rho^{|j-i|}. \quad (\text{S3.22})$$

This is actually one of the few cases where an analytic expression for the eigenvectors is known, although it is rather complex (Rao & Yip, 1990; Ray & Driver, 1970). Interestingly, one can use this to show that for AR(1) processes, the discrete cosine transform DCT-II becomes equivalent to the ‘principal component transform’ (Karhunen-Loève transform) as $\rho \rightarrow 1$ (Rao & Yip, 1990, section 3.3.2).

High-dimensional random walk In the limit $\rho \rightarrow 1$ one obtains a random walk. Antognini and Sohl-Dickstein (2018) analyse the principal components of high-dimensional random walks. We briefly summarise their results. Consider a random walk in \mathbb{R}^M with N steps given by

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathbf{r}_n \quad (\text{S3.23})$$

where \mathbf{r}_n is a random step drawn from a probability distribution with zero mean and a finite, normalized covariance matrix. We start from $\mathbf{x}_0 = \mathbf{0}$ in \mathbb{R}^M .

We can express all this as matrix multiplications. Collect the points \mathbf{x}_n and steps \mathbf{r}_n as the rows of the $N \times M$ matrices \mathbf{X} and \mathbf{R} respectively. Let \mathbf{W} be a $N \times N$ matrix with 1’s on the diagonal, -1 ’s on the

subdiagonal and zeros elsewhere. This implements the walking mechanism in the sense that $\mathbf{W}\mathbf{X} = \mathbf{R}$, hence

$$\mathbf{X} = \mathbf{W}^{-1}\mathbf{R}. \quad (\text{S3.24})$$

To compute the covariance matrix \mathbf{S} we need the centered datapoints $\hat{x}_n = x_n - \bar{x}_n$. The centering operation be conveniently expressed as multiplication by the $N \times N$ centering matrix $\mathbf{C} = \mathbf{I} - \frac{1}{M}\mathbf{J}$, where \mathbf{J} is the all-ones matrix. This gives

$$\hat{\mathbf{X}} = \mathbf{C}\mathbf{X} = \mathbf{C}\mathbf{W}^{-1}\mathbf{R} \quad (\text{S3.25})$$

and allows us to express the covariance matrix as $\mathbf{S} = \frac{1}{N}\hat{\mathbf{X}}^T\hat{\mathbf{X}}$. Instead of finding the eigenvectors of $\hat{\mathbf{X}}^T\hat{\mathbf{X}}$, we can look for those of $\hat{\mathbf{X}}\hat{\mathbf{X}}^T$. After all, if \mathbf{u} is an eigenvector for $\hat{\mathbf{X}}^T\hat{\mathbf{X}}$ with nonzero eigenvalue λ , then $\mathbf{v} = \hat{\mathbf{X}}\mathbf{u}$ is the corresponding eigenvector for $\hat{\mathbf{X}}\hat{\mathbf{X}}^T$.

Putting all this together, Antognini and Sohl-Dickstein (2018) look for the eigenvalues and eigenvectors of

$$\hat{\mathbf{X}}\hat{\mathbf{X}}^T = \mathbf{C}\mathbf{W}^{-1}\mathbf{R}\mathbf{R}^T\mathbf{W}^{-T}\mathbf{C} \quad (\text{S3.26})$$

where we used symmetry of \mathbf{C} . Note that this matrix contains the covariance between timesteps, rather than dimensions. They observe that in the limit of infinite dimensionality $M \rightarrow \infty$, we have that $\mathbf{R}\mathbf{R}^T$ tends to the $N \times N$ identity matrix. This allows us to simplify (S3.26) to

$$\hat{\mathbf{X}}\hat{\mathbf{X}}^T = \mathbf{C}\mathbf{W}^{-1}\mathbf{W}^{-T}\mathbf{C}. \quad (\text{S3.27})$$

Since \mathbf{W} is a so called banded Toeplitz matrix, and \mathbf{C} is circulant, the whole expression can be shown to be asymptotically equivalent to a circulant matrix, meaning that the eigenvectors are cosines. This analysis can be related to melodic contours, when we consider a collection of M contours of length N as one high-dimensional walk through \mathbb{R}^M .

Bibliography

- Antognini, J. M., & Sohl-Dickstein, J. (2018). PCA of high dimensional random walks with comparison to neural network training. *32nd Conference on Neural Information Processing Systems*.
- Gray, R. M. (2006). Toeplitz and Circulant Matrices: A Review. *Foundations and Trends in Communications and Information Theory*, 2(3), 155–239. <https://doi.org/10/bxhj7f>
- Jolliffe, I. (2002). *Principal Component Analysis* (2nd ed.). Springer.
- Rao, K. R., & Yip, P. C. (1990). *Discrete cosine transform: Algorithms, advantages, applications*. Academic Press.
- Ray, W. D., & Driver, R. M. (1970). Further Decomposition of the Karhunen-Loève Series Representation of a Stationary Random Process. *IEEE Transactions on Information Theory*, 16(6), 663–668.