Chapter 4: Visualization

1. Diagrams
2. Principal Component Analysis
3. Multi Dimensional Scaling
4. Sammon Mapping
5. Auto–Encoder
6. Histograms
7. Spectral Analysis

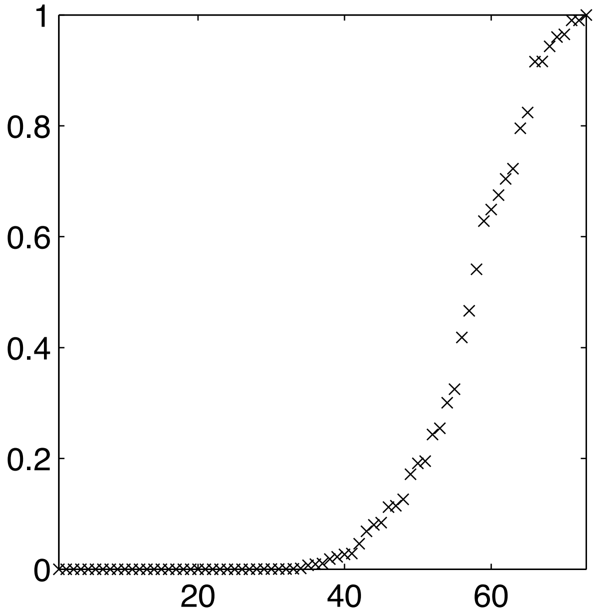
Diagrams

1 feature

dimension $p = 1$

$X = \{x_k\}$

diagram $x_1, \ldots, x_n$

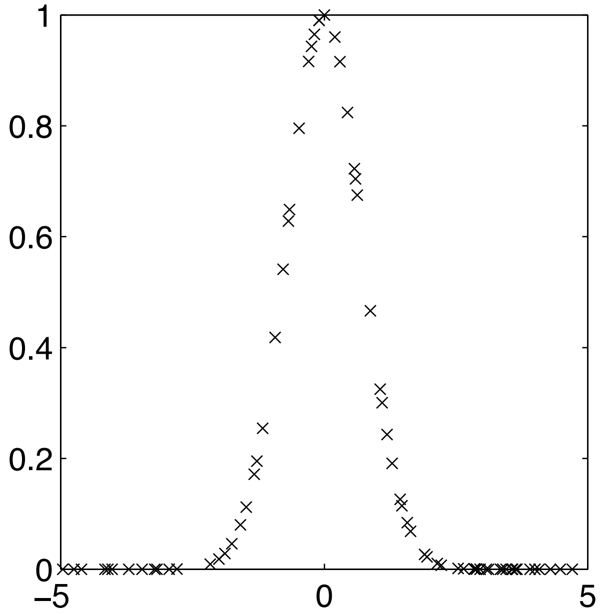


2 features (correlation between 2 features)

dimension $p = 2$

$X = \{(x_k, y_k)\}$

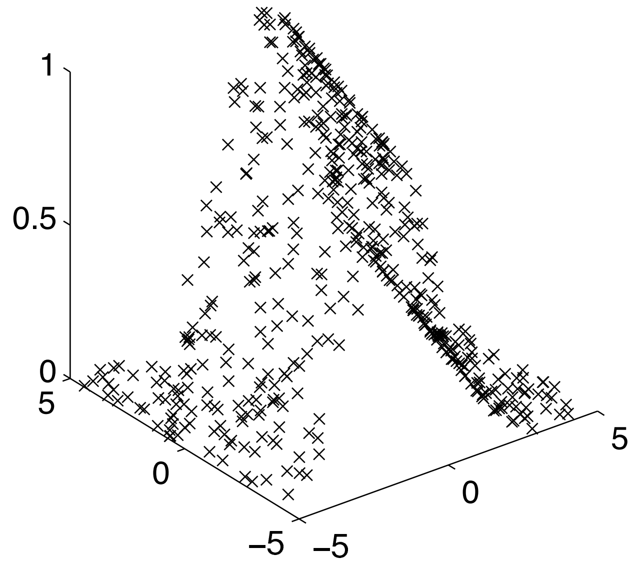
scatter diagram $x(y)$



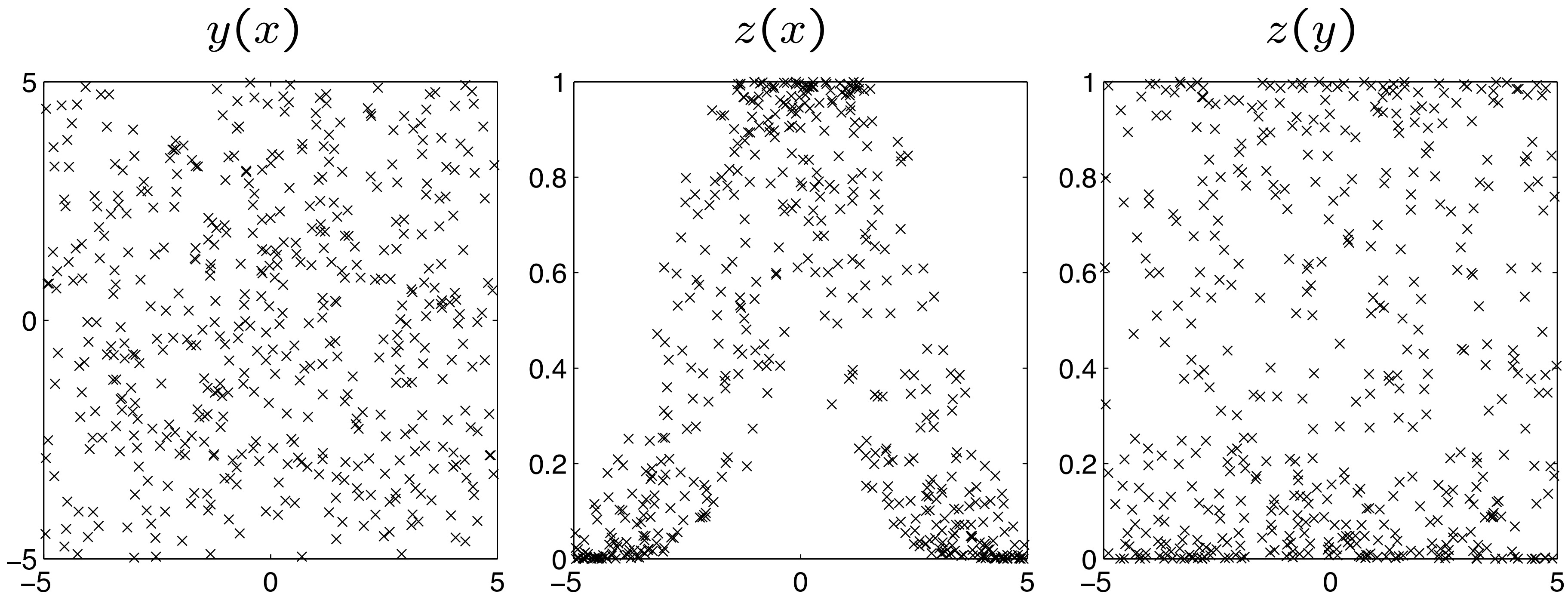
dimension $p = 3$

$X = \{(x_k, y_k, z_k)\}$

3D scatter diagram $z(x, y)$



Projection



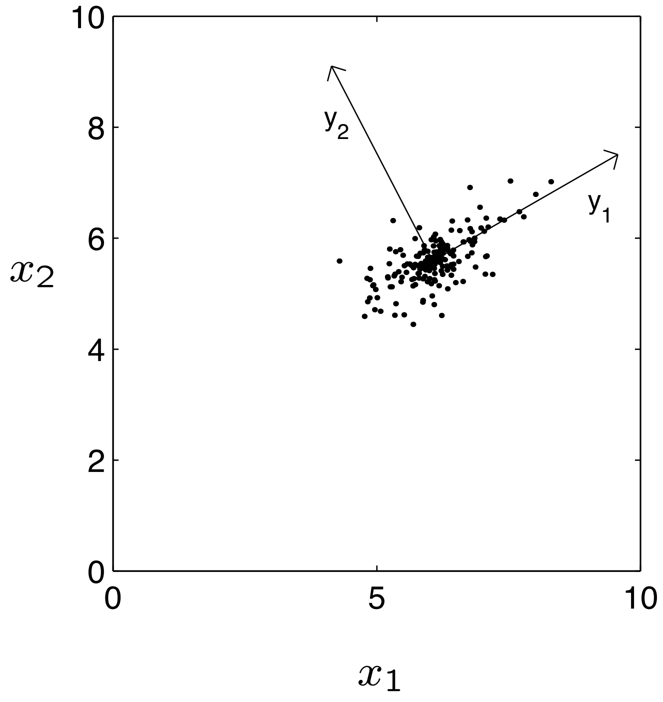
Prof. Dr. Thomas A. Runkler

Principal Component Analysis

Different names:

- principal component analysis (PCA)
- Karhunen-Loève transform
- singular value decomposition (SVD)
- eigenvector projection
- Hotelling transform
- proper orthogonal decomposition

rotate the data to get the max. variance



Prof. Dr. Thomas A. Runkler

Principal Component Analysis

- coordinate transformation: translation and rotation

  transformation $\mathbb{R}^p \rightarrow \mathbb{R}^q$: $\quad y_k = (x_k - \bar{x}) \cdot E$

  inverse transformation $\mathbb{R}^q \rightarrow \mathbb{R}^p$: $\quad x_k = y_k \cdot E^T + \bar{x}$

- mean

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

- determine rotation matrix E by maximizing the variance of Y

Principal Component Analysis

- variance in Y

$$
\begin{aligned}
v_y &= \frac{1}{n-1}\sum_{k=1}^{n} y_k^T y_k \\
&= \frac{1}{n-1}\sum_{k=1}^{n} ((x_k-\bar{x})\cdot E)^T\cdot((x_k-\bar{x})\cdot E) \\
&= \frac{1}{n-1}\sum_{k=1}^{n} E^T\cdot(x_k-\bar{x})^T\cdot(x_k-\bar{x})\cdot E \\
&= E^T\left(\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\bar{x})^T\cdot(x_k-\bar{x})\right)\cdot E \\
&= E^T\cdot C\cdot E
\end{aligned}
$$

$y_k = (x_k - \bar{x}) \cdot E$

$(AB)^T = B^T A^T$

E^T and E are constants

$\rightarrow$ maximize this equation

- covariance matrix of X

$$
c_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k^{(i)}-\bar{x}^{(i)})(x_k^{(j)}-\bar{x}^{(j)})
$$

Principal Component Analysis

- constraint: rotation, no dilation

$$E^T \cdot E = I$$

- Lagrange optimization

$$\begin{aligned} L &= E^T C E - \lambda(E^T E - I) \\ \frac{\partial L}{\partial E} = 0 &\Rightarrow CE - \lambda E + E^T C - \lambda E^T = 0 \\ &\Rightarrow CE = \lambda E \text{ (eigenproblem)} \\ \text{solution} \quad & (v_1, \ldots, v_p, \lambda_1, \ldots, \lambda_p) = \text{eig}(C) \end{aligned}$$

- solution using homogeneous equation system

$$(C - \lambda I) \cdot E = 0$$

Principal Component Analysis

- E is matrix of eigenvectors of C

$$E = (v_1, \ldots, v_q)$$

- variances in Y are eigenvalues of C

$$CE = \lambda E \quad \Leftrightarrow \quad \lambda = E^T C E = v_y$$

- suitable dimensionality q

$$\sum_{i=1}^{q} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i \geq 95\%$$

- transformation error

$$e = \frac{1}{n} \sum_{k=1}^{n} (x_k - x'_k)^2 = \frac{n-1}{n} \sum_{i=q+1}^{p} \lambda_i$$

$\Rightarrow$ PCA yields projection with minimal quadratic error

= maximize the variance

Example Principal Component Analysis

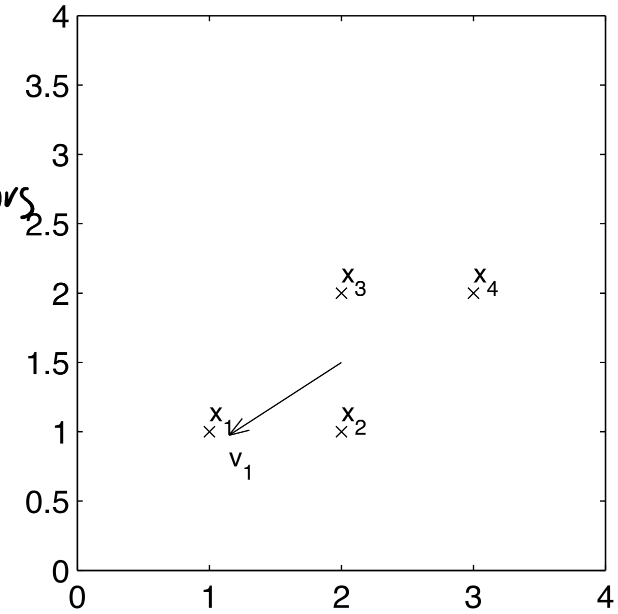
PCA

$$
\begin{aligned}
X &= \{(1,1),(2,1),(2,2),(3,2)\} \\
\bar{x} &= \frac{1}{2}\cdot(4,3) \\
C &= \frac{1}{3}\cdot \begin{array}{c c} & \begin{array}{cc} x & y \end{array} \\ \begin{array}{c} x \\ y \end{array} & \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \end{array} \longrightarrow \\
\lambda_1 &= 0.8727 \\
\lambda_2 &= 0.1273 \\
v_1 &= \begin{pmatrix} -0.85065 \\ -0.52573 \end{pmatrix} \\
v_2 &= \begin{pmatrix} 0.52573 \\ -0.85065 \end{pmatrix}
\end{aligned}
$$

find eigenvalues
and its eigenvectors

$(\lambda_1 - v_1)$

$(\lambda_2 - v_2)$



Example Principal Component Analysis

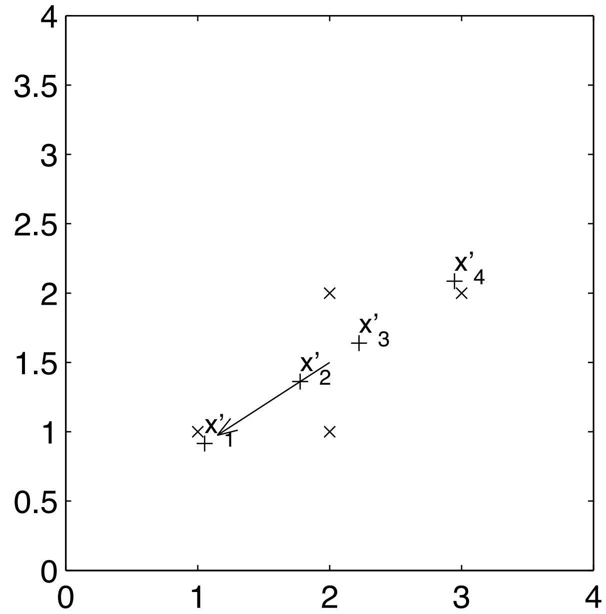
projection onto first axis:

rotation matrix

$$E = v_1 = \begin{pmatrix} -0.85065 \\ -0.52573 \end{pmatrix}$$

$$Y = (1.1135, 0.2629, \\ -0.2629, -1.1135)$$

$$X' = \{(1.0528, 0.91459), \\ (1.7764, 1.3618), \\ (2.2236, 1.6382), \\ (2.9472, 2.0854)\}$$



Prof. Dr. Thomas A. Runkler

Multi Dimensional Scaling

equivalent to PCA

- eigendecomposition of the positive semi–definite matrix XX^T

$$XX^T = Q\Lambda Q^T = (Q\sqrt{\Lambda}^T) \cdot (\sqrt{\Lambda}Q^T) = (Q\sqrt{\Lambda}^T) \cdot (Q\sqrt{\Lambda}^T)^T$$

  $Q = (v_1, \ldots, v_p)$: matrix of eigenvectors of XX^T, usually $p < n$

  Λ: diagonal matrix of eigenvalues of XX^T

- estimate for X

$$Y = Q\sqrt{\Lambda}^T$$

- lower dimensional projections $Y \subset R^q$, $q < p$
  - use only first q dimensions
  - scale eigenvectors (square norms match eigenvalues)

MDS of a Distance Matrix D

- assume $\tilde{X}$ and choose anchor point $\tilde{x}_a$, $a \in \{1, \ldots, n\}$
- transform $\tilde{X}$ (origin 0) to X (origin $\tilde{x}_a$)

$$x_k = \tilde{x}_k - \tilde{x}_a$$

- difference vectors are invariant

$$\tilde{x}_i - \tilde{x}_j = x_i - x_j$$

- scalar product of each side with itself

$$(\tilde{x}_i - \tilde{x}_j)(\tilde{x}_i - \tilde{x}_j)^T = (x_i - x_j)(x_i - x_j)^T$$

$$\Rightarrow d_{ij}^2 = x_i x_i^T - 2 x_i x_j^T + x_j x_j^T = d_{ia}^2 - 2 x_i x_j^T + d_{ja}^2$$

$$\Rightarrow x_i x_j^T = (d_{ia}^2 + d_{ja}^2 - d_{ij}^2)/2$$

- XX^T can be computed from D

Example Multi Dimensional Scaling

$$X' = \{(1,1),(2,1),(2,2),(3,2)\}$$

$$X = \left\{(-1,-\frac{1}{2}),(0,-\frac{1}{2}),(0,\frac{1}{2}),(1,\frac{1}{2})\right\} \quad \text{(mean subtracted)}$$

$$XX^T = \frac{1}{4}\begin{pmatrix} 5 & 1 & -1 & -5 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -5 & -1 & 1 & 5 \end{pmatrix}$$

eigenvector/eigenvalue solver

$$v_1 \approx \begin{pmatrix} -0.6882 \\ -0.1625 \\ 0.1625 \\ 0.6882 \end{pmatrix}, \quad \lambda_1 \approx 2.618$$

$\sqrt{\lambda_1}$

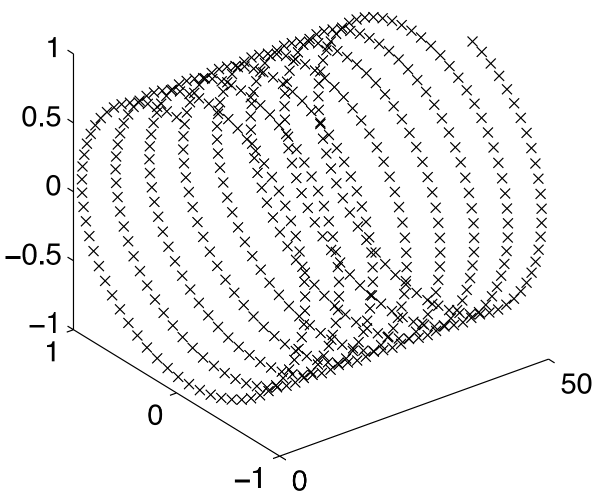
$$\Rightarrow Y \approx \begin{pmatrix} -0.6882 \\ -0.1625 \\ 0.1625 \\ 0.6882 \end{pmatrix} \sqrt{2.618} \approx \begin{pmatrix} -1.1135 \\ -0.2629 \\ 0.2629 \\ 1.1135 \end{pmatrix}$$

Prof. Dr. Thomas A. Runkler

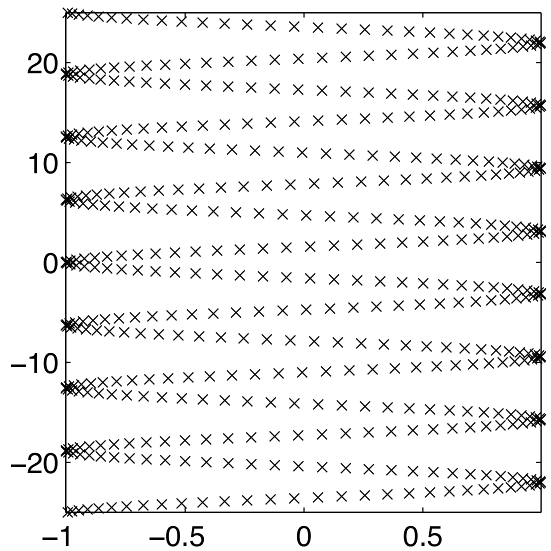
Example MDS: Helix

$$X = \{(t, \sin t, \cos t)^T \mid t \in \{0, 0.1, 0.2, \ldots, 50\}\}$$

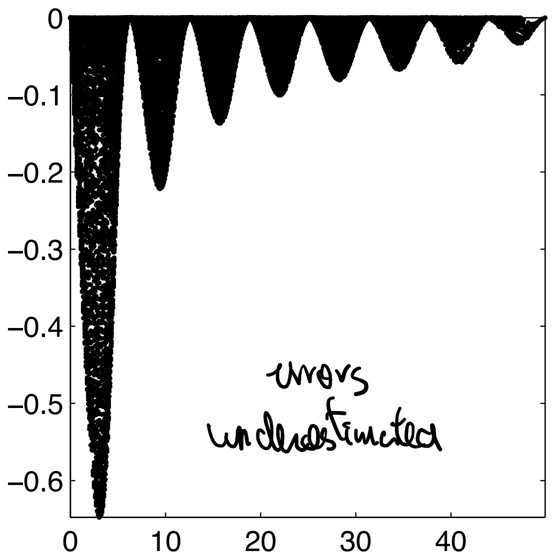
original data



MDS projection



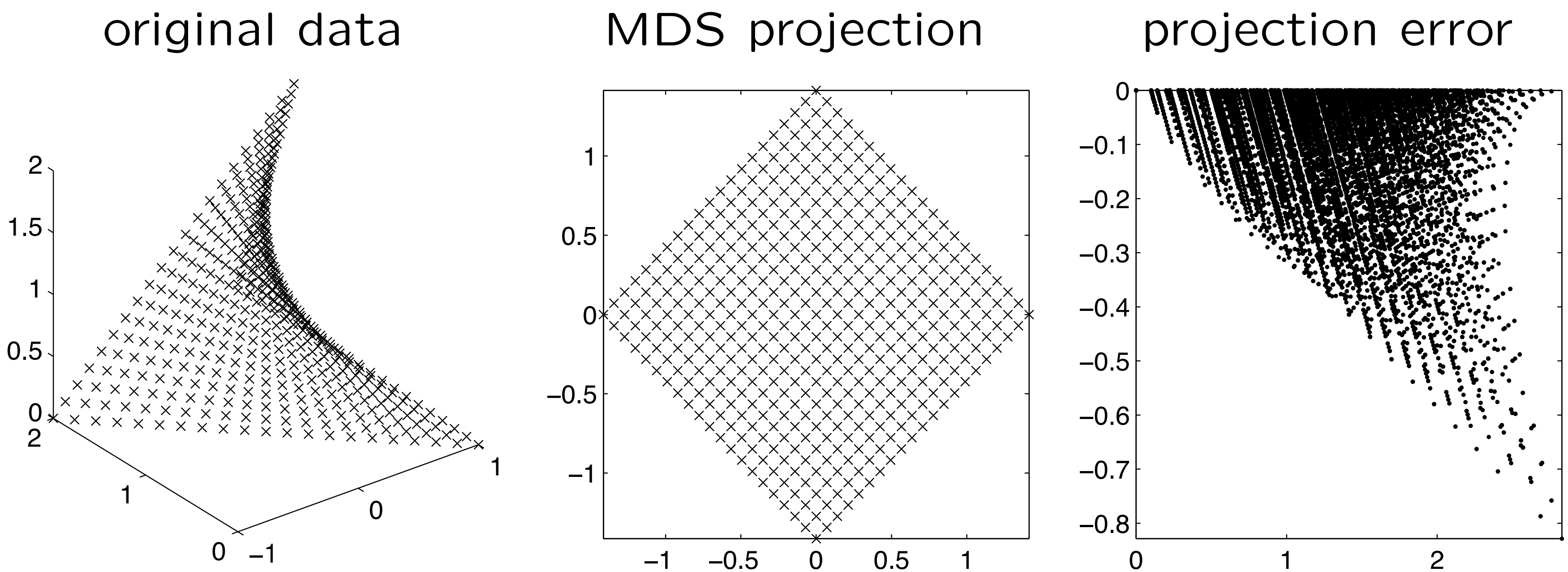
projection error



Prof. Dr. Thomas A. Runkler

Example MDS: Bent Square

$$X = \{((t_1 - 1) \cdot (t_2 - 1), t_1, t_2)^T \mid t_1, t_2 \in \{0, 0.1, 0.2, \ldots, 2\}\}$$



Prof. Dr. Thomas A. Runkler

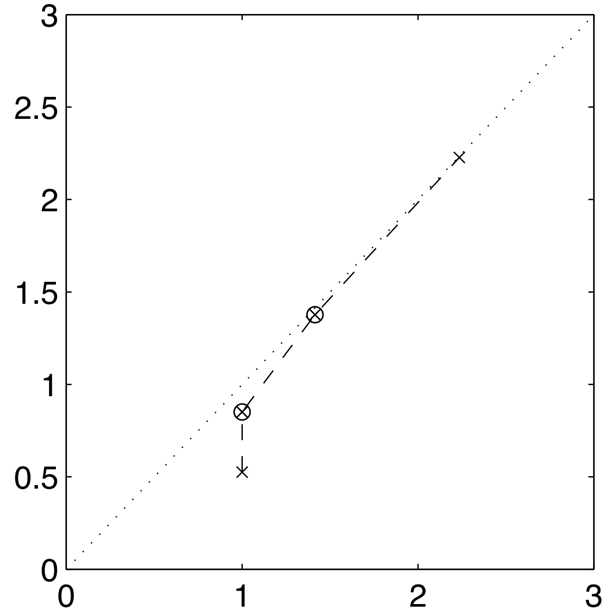
Shepard Diagram

to evaluate the performance of the projection

$$D^x \approx \begin{pmatrix} 0 & 1 & 1.4142 & 2.2361 \\ 1 & 0 & 1 & 1.4142 \\ 1.4142 & 1 & 0 & 1 \\ 2.2361 & 1.4142 & 1 & 0 \end{pmatrix}$$

smaller (underestimate)

$$D^y \approx \begin{pmatrix} 0 & 0.8507 & 1.3764 & 2.2270 \\ 0.8507 & 0 & 0.5257 & 1.3764 \\ 1.3764 & 0.5257 & 0 & 0.8507 \\ 2.2270 & 1.3764 & 0.8507 & 0 \end{pmatrix}$$



- alternative criteria for multidimensional scaling
  - strict monotonicity (Torgerson)
  - points close to main diagonal (Sammon)

(non linear)

Sammon Mapping

- original data $X = \{x_1, \ldots, x_n\} \subset \mathrm{IR}^p$
- distance matrix $D^x_{ij} = \|x_i - x_j\|$
- transformed data $Y = \{y_1, \ldots, y_n\} \subset \mathrm{IR}^q$
- distance matrix $D^y_{ij} = \|y_i - y_j\|$

These 2 distances should be similar (close)

- wanted:

  (nonlinear) mapping $X \to Y$,

  so that $D^x_{ij} \approx D^y_{ij}$ $\forall i, j = 1, \ldots, n$

  minimizing the error

- or:

  given relational data D^x,

  wanted corresponding real valued object data Y,

  so that $D^x_{ij} \approx D^y_{ij}$ $\forall i, j = 1, \ldots, n$

Sammon Mapping

- error measures:

$$E_1 = \frac{1}{\sum\limits_{i=1}^{n}\sum\limits_{j=i+1}^{n}\left(D_{ij}^x\right)^2} \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(D_{ij}^y - D_{ij}^x\right)^2$$

$$E_2 = \sum_{i=1}^{n}\sum_{j=i+1}^{n}\left(\frac{D_{ij}^y - D_{ij}^x}{D_{ij}^x}\right)^2$$

$$E_3 = \frac{1}{\sum\limits_{i=1}^{n}\sum\limits_{j=i+1}^{n} D_{ij}^x} \sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{\left(D_{ij}^y - D_{ij}^x\right)^2}{D_{ij}^x}$$

- E_1 minimizes the global absolute error (not very good)
- E_2 minimizes the relative local error (also not very good)
- E_3 is a compromise between E_1 and E_2 (best choice) *

Numerical Optimization

- gradient descent
  1. input X, α, initialize Y
  2. repeat

     derivative respect y_k

     $$y_k := y_k - \alpha \frac{\partial E}{\partial y_k}, \qquad k = 1, \ldots, n$$

     until termination
  3. output Y
- Newton's method
  1. input X, initialize Y
  2. repeat

     derivative

     first second

     $$y_k := y_k - \left(\frac{\partial E}{\partial y_k} \right) \Big/ \left(\frac{\partial^2 E}{\partial y_k^2} \right), \qquad k = 1, \ldots, n$$

     until termination
  3. output Y

Prof. Dr. Thomas A. Runkler

Derivatives of the Sammon Function

- preliminary note

$$\frac{\partial D_{ij}^y}{\partial y_k} = \frac{\partial}{\partial y_k}\|y_i - y_j\| = \begin{cases} \frac{y_k - y_j}{D_{kj}^y} & \text{if } i = k \\ 0 & \text{otherwise} \end{cases} \quad (i \neq k \text{ and } j \neq k)$$

- first derivative

$$\frac{\partial E_3}{\partial y_k} = \frac{2}{\sum\limits_{i=1}^{n}\sum\limits_{j=i+1}^{n} d_{ij}^x} \sum_{\substack{j=1\\ j\neq k}}^{n} \left(\frac{1}{d_{kj}^x} - \frac{1}{d_{kj}^y}\right)\left(y_k - y_j\right)$$

- second derivative

$$\frac{\partial^2 E_3}{\partial y_k^2} = \frac{2}{\sum\limits_{i=1}^{n}\sum\limits_{j=i+1}^{n} d_{ij}^x} \sum_{\substack{j=1\\ j\neq k}}^{n} \left(\frac{1}{d_{kj}^x} - \frac{1}{d_{kj}^y} - \frac{(y_k - y_j)^2}{(d_{kj}^y)^3}\right)$$

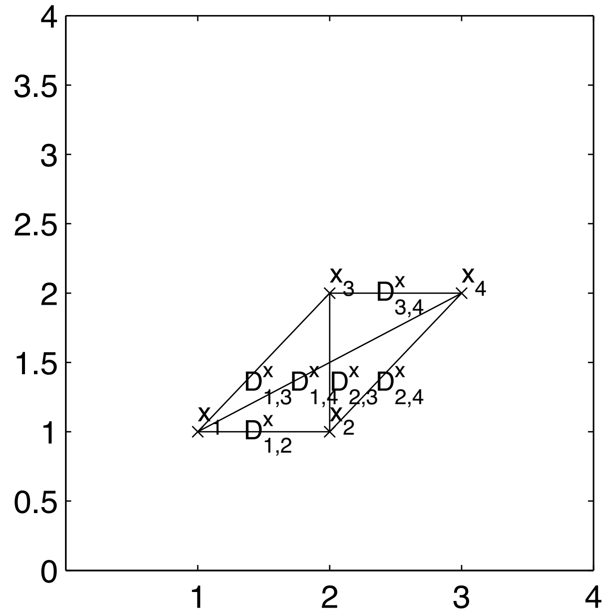
Example Sammon Mapping

$$X = \{(1,1),(2,1),(2,2),(3,2)\}$$

$$D^x = \begin{pmatrix} 0 & 1 & \sqrt{2} & \sqrt{5} \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ \sqrt{5} & \sqrt{2} & 1 & 0 \end{pmatrix}$$

$$Y = \{1,2,3,4\}$$

$$D^y = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix}$$



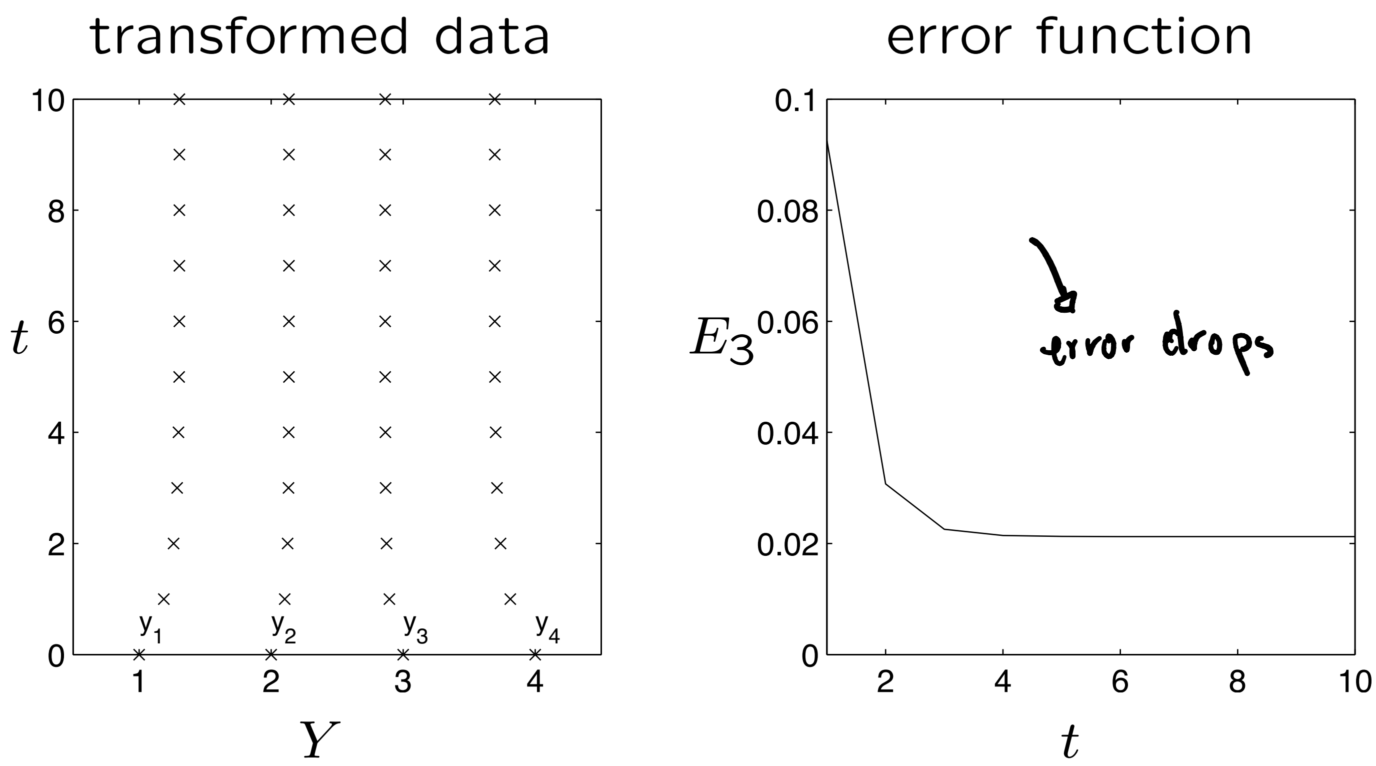
Prof. Dr. Thomas A. Runkler

Example Sammon Mapping

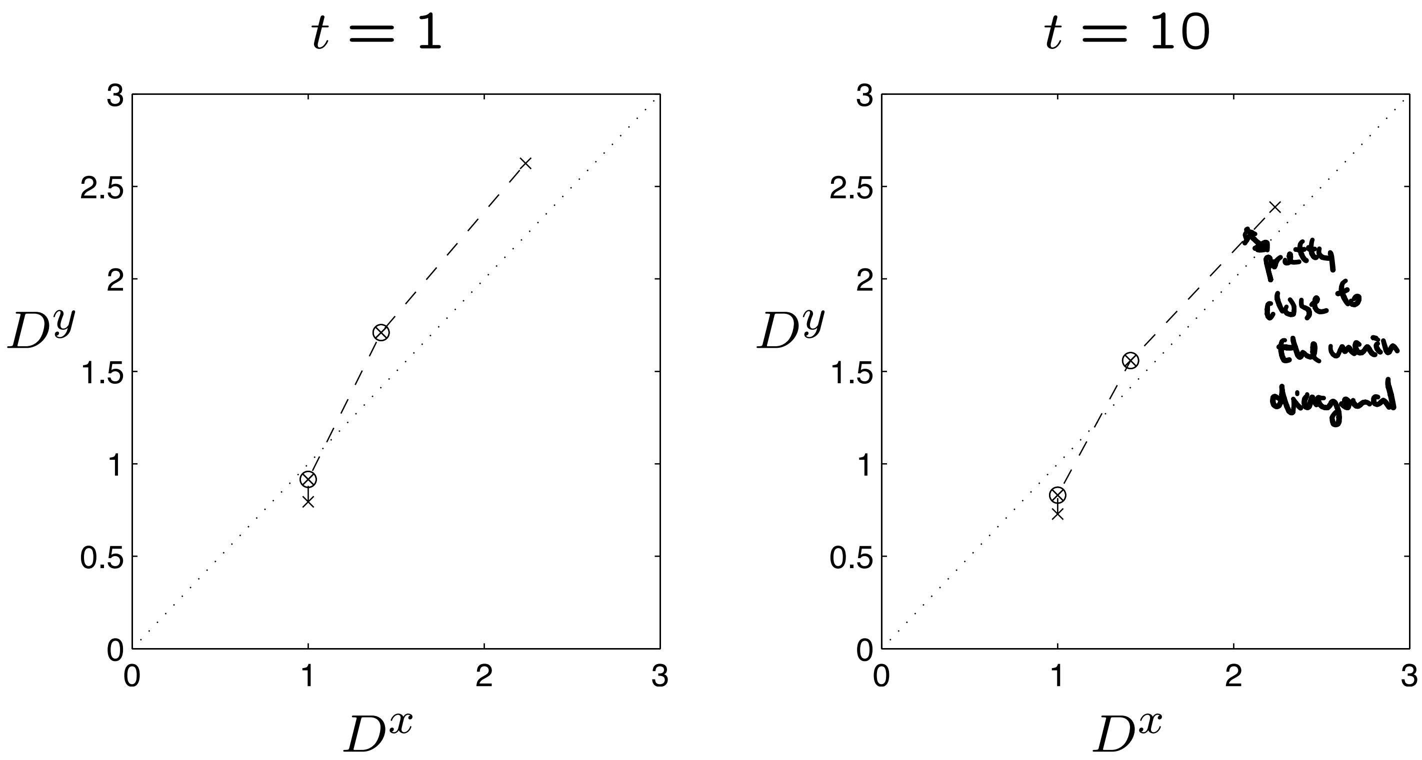
$$
\begin{aligned}
E_3 &= \frac{1}{3\cdot 1+2\cdot\sqrt{2}+\sqrt{5}}\cdot\left(2\cdot\frac{\left(2-\sqrt{2}\right)^2}{\sqrt{2}}+\frac{\left(3-\sqrt{5}\right)^2}{\sqrt{5}}\right)=0.0925\\
\frac{\partial E_3}{\partial y_1} &= \frac{2}{3\cdot 1+2\cdot\sqrt{2}+\sqrt{5}}\cdot\left(-\frac{2-\sqrt{2}}{\sqrt{2}}-\frac{3-\sqrt{5}}{\sqrt{5}}\right)=-0.1875\\
\frac{\partial E_3}{\partial y_2} &= \frac{2}{3\cdot 1+2\cdot\sqrt{2}+\sqrt{5}}\cdot\left(-\frac{2-\sqrt{2}}{\sqrt{2}}\right)=-0.1027\\
\frac{\partial E_3}{\partial y_3} &= \frac{2}{3\cdot 1+2\cdot\sqrt{2}+\sqrt{5}}\cdot\left(\frac{2-\sqrt{2}}{\sqrt{2}}\right)=0.1027\\
\frac{\partial E_3}{\partial y_4} &= \frac{2}{3\cdot 1+2\cdot\sqrt{2}+\sqrt{5}}\cdot\left(\frac{3-\sqrt{5}}{\sqrt{5}}+\frac{2-\sqrt{2}}{\sqrt{2}}\right)=0.1875\\
y &\leftarrow (1.1875,\ 2.1027,\ 2.8973,\ 3.8125)\ \text{new values of } Y
\end{aligned}
$$

Prof. Dr. Thomas A. Runkler

Example Sammon Mapping



Shepard Diagrams

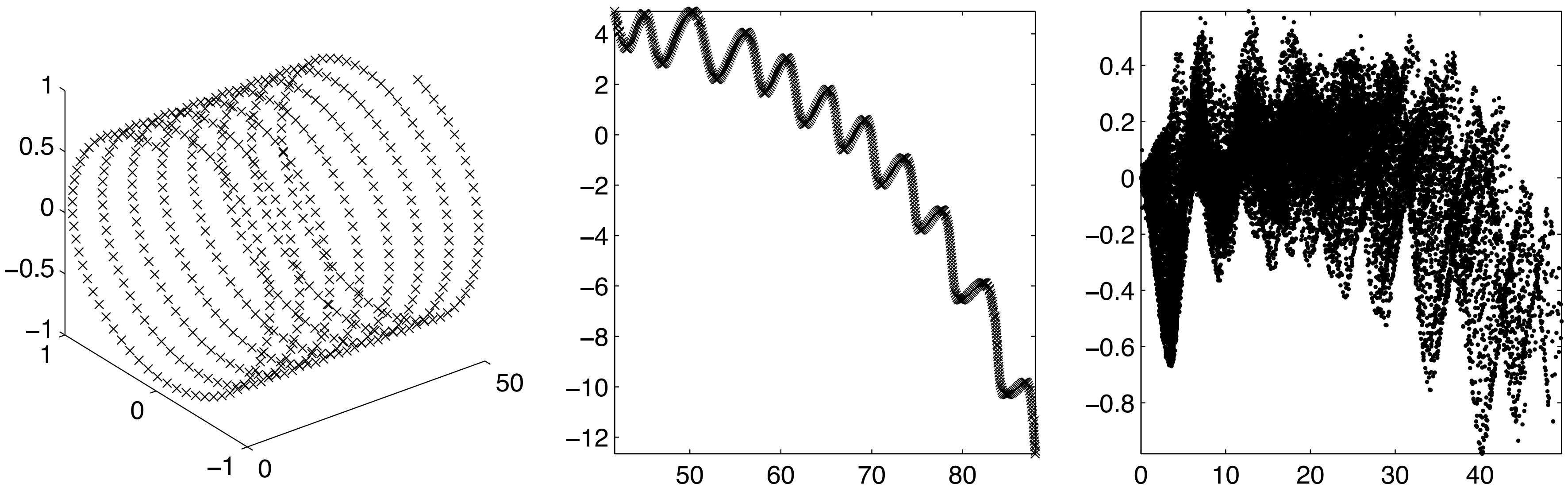


Prof. Dr. Thomas A. Runkler

Example Sammon Mapping: Helix

$$X = \{(t, \sin t, \cos t)^T \mid t \in \{0, 0.1, 0.2, \ldots, 50\}\}$$

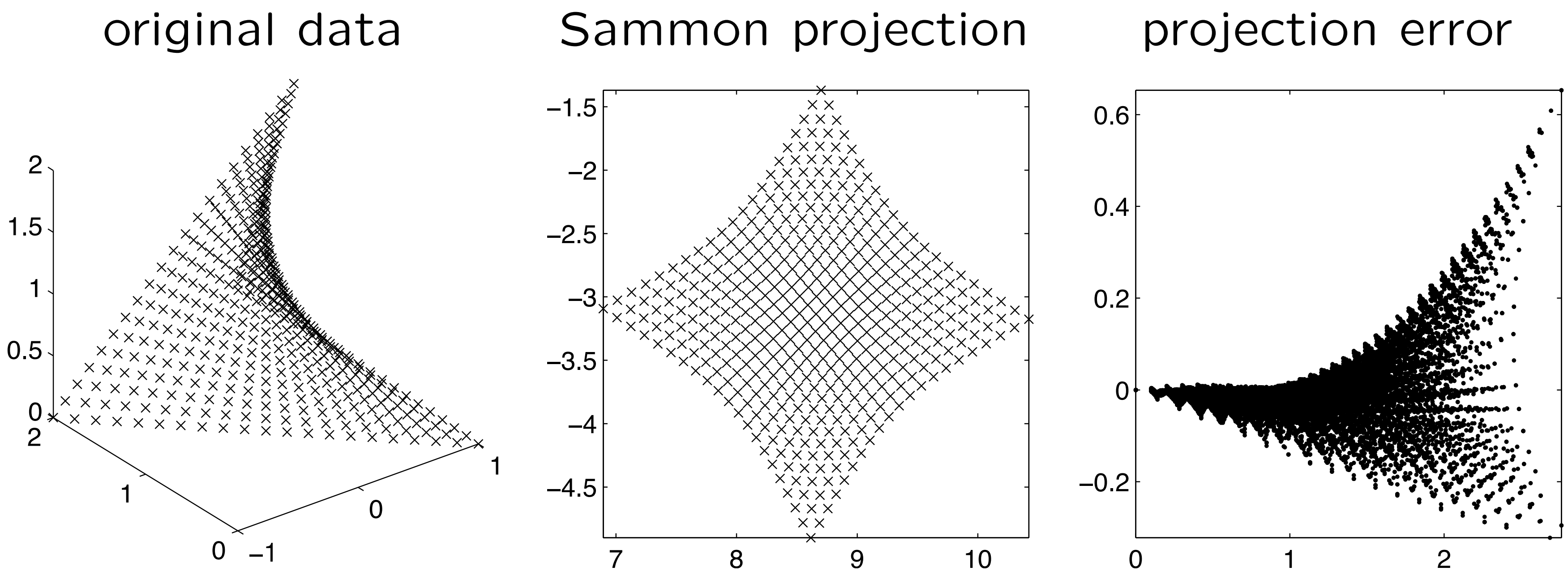
original data　　Sammon projection　　projection error



Prof. Dr. Thomas A. Runkler

Example Sammon Mapping: Bent Square

$$X = \{((t_1 - 1) \cdot (t_2 - 1), t_1, t_2)^T \mid t_1, t_2 \in \{0, 0.1, 0.2, \ldots, 2\}\}$$



Prof. Dr. Thomas A. Runkler

Auto–Encoder → core of I2DL

2 mappings

- find $f : R^p \to R^q$ and $g : R^q \to R^p$

$$\begin{aligned} y_k &= f(x_k) \\ x_k &\approx g(y_k) \end{aligned}$$

- find $g \circ f$ by regression   dual-regression

$$x_k \approx g \circ f(x_k) = g(f(x_k))$$

- use f for mapping

$$y_k = f(x_k)$$

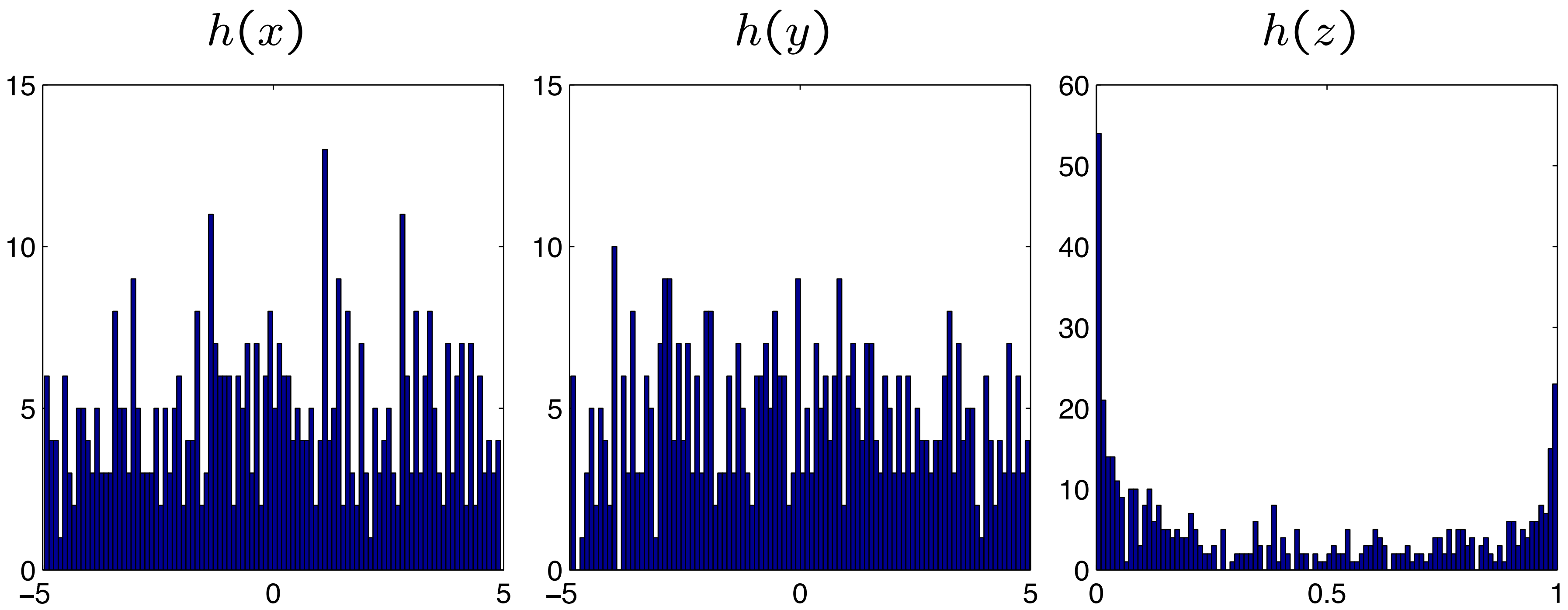
- suitable regression models will be presented in chapter 6
- neural network auto–encoders are widely used as layers of deep learning architectures

visualizing data

Histogram

$$h_k(X) = |\{\xi \in X \mid \xi_k \leq \xi < \xi_{k+1}\}|, \quad k = 1, \ldots, m$$

$$\Delta x = (\max X - \min X)/m, \quad \xi_k = \min X + (k-1) \cdot \Delta x$$



Number of Histogram Intervals

- Sturgess: number of data

$$m = 1 + \log_2 n$$

- Scott: Gaussian distribution

$$m = \frac{3.49 \cdot s}{\sqrt[3]{n}}$$

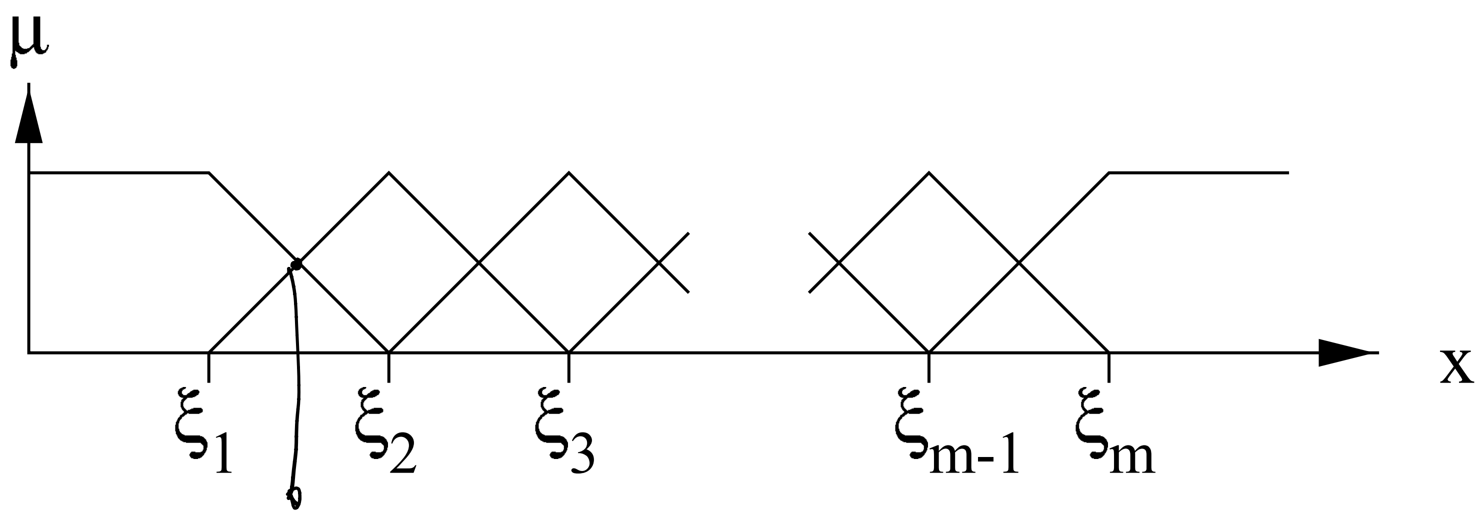
standard deviation

- Freedman and Diaconis: middle 50% quantile

$$m = \frac{2 \cdot (Q_{75\%} - Q_{25\%})}{\sqrt[3]{n}}$$

$$|\{x \in X \mid x \leq Q_\varphi\}| = \varphi \cdot n$$

Fuzzy Histogram



50% ξ_1

50% ξ_2

$$\tilde{h}_k(X) = \sum_{x \in X} \mu_k(x)$$

Spectral Analysis (periodic components)

- Fourier theorem

$$
\begin{aligned}
f(x) &= \int_0^{\infty} \left(a(y)\,\cos xy + b(y)\,\sin xy\right)\,dy \quad \text{with} \\
a(y) &= \frac{1}{\pi}\int_{-\infty}^{\infty} f(u)\,\cos yu\,du \\
b(y) &= \frac{1}{\pi}\int_{-\infty}^{\infty} f(u)\,\sin yu\,du
\end{aligned}
$$

Spectral Analysis

- Fourier cosine transform

$$F_c(y) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \cos xy \, dx$$ (forward)

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty F_c(y) \cos xy \, dy$$ (backward)

- Fourier sine transform

$$F_s(y) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x) \sin xy \, dx$$ (forward)

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty F_s(y) \sin xy \, dy$$ (backward)

Spectral Analysis

- discretization: $x = k \cdot T$, $y = l \cdot \omega$
- discrete Fourier cosine transform

$$y_l^c = \frac{2}{n} \sum_{k=1}^{n} x_k \cos kl\omega T$$

$$x_k' = \frac{n\omega T}{\pi} \sum_{l=1}^{m} y_l^c \cos kl\omega T$$

- discrete Fourier sine transform

$$y_l^s = \frac{2}{n} \sum_{k=1}^{n} x_k \sin kl\omega T$$

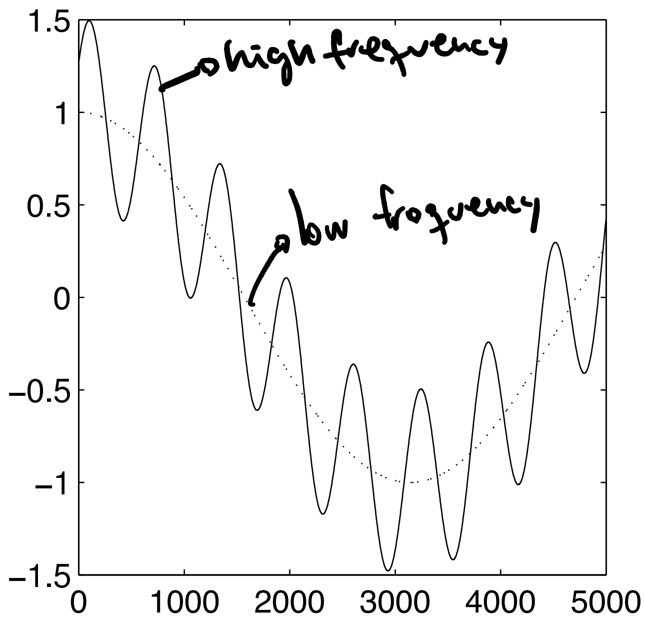
$$x_k' = \frac{n\omega T}{\pi} \sum_{l=1}^{m} y_l^s \sin kl\omega T$$

Example Cosine Transform

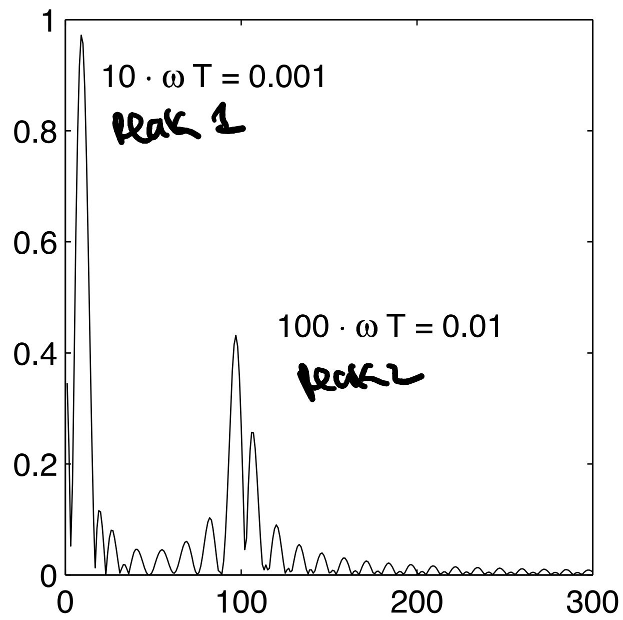
$X = \{(i, \cos(\overset{\text{peak 1}}{0.001} \cdot i) + 0.5 \cdot \cos(\overset{\text{peak 2}}{0.01} \cdot i - 1)) \mid i \in \{1, \ldots, 5000\}\}$

$m = 300$ spectral values, $\omega T = 10^{-4}$

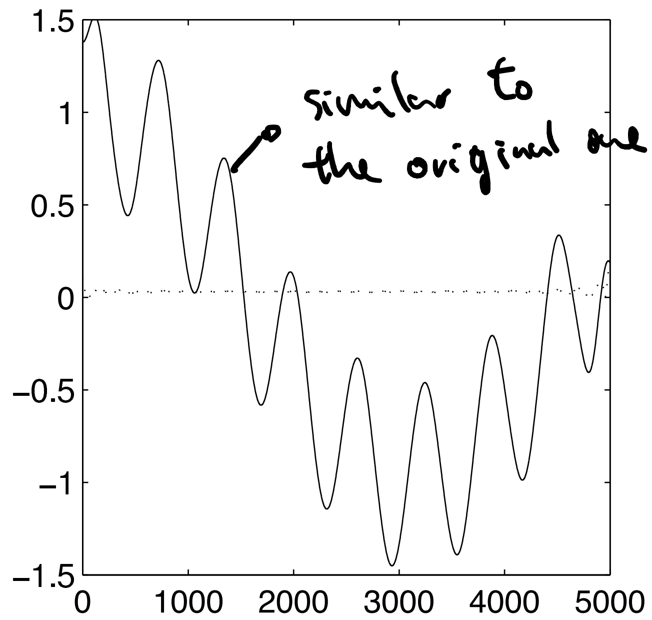
original data
$X = \{x_1, \ldots, x_{5000}\}$



cosine spectrum
$|Y^c| = \{|y_1^c|, \ldots, |y_{300}^c|\}$



inverse transform
$X' = \{x'_1, \ldots, x'_{5000}\}$

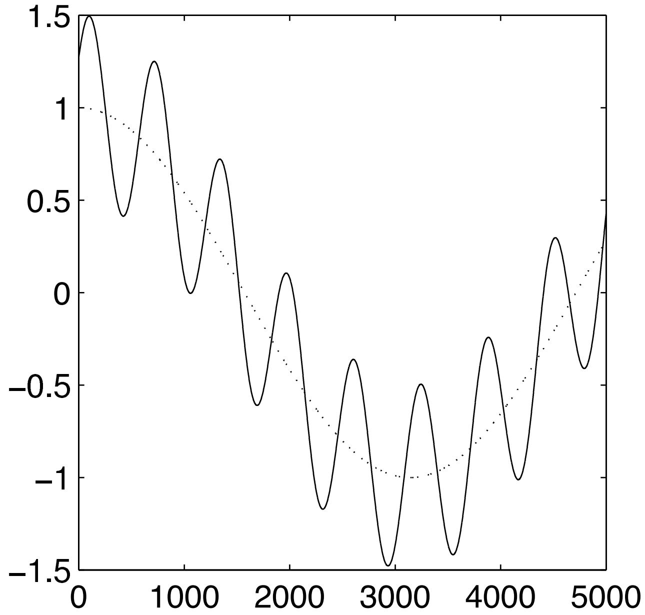


Example Sine Transform

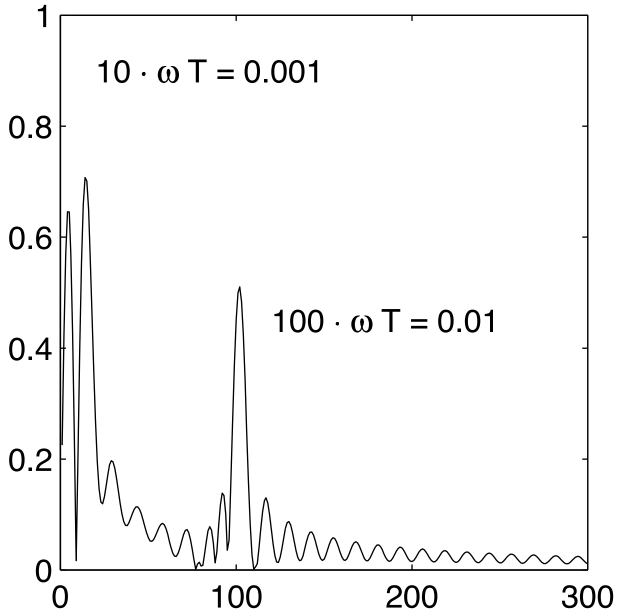
$$X = \{(i, \cos(0.001 \cdot i) + 0.5 \cdot \cos(0.01 \cdot i - 1)) \mid i \in \{1, \ldots, 5000\}\}$$

$m = 300$ spectral values, $\omega T = 10^{-4}$

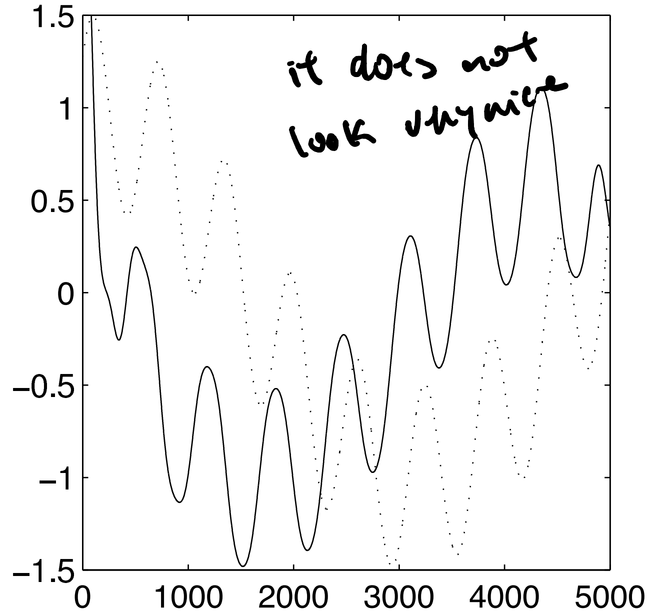
original data
$X = \{x_1, \ldots, x_{5000}\}$



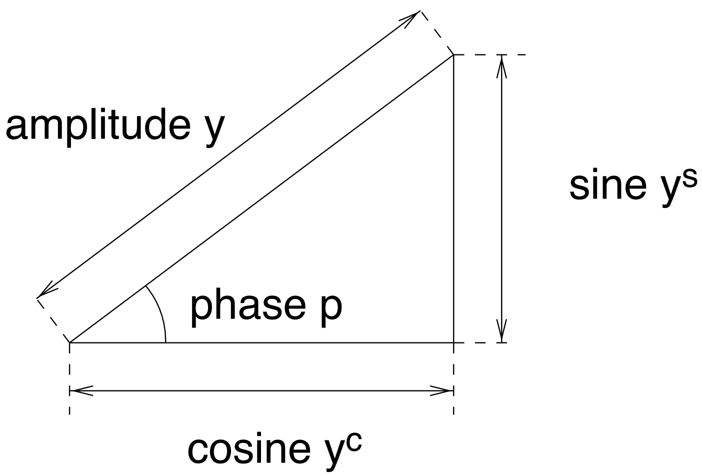
sine spectrum
$|Y^s| = \{|y^s_1|, \ldots, |y^s_{300}|\}$



inverse transform
$X' = \{x'_1, \ldots, x'_{5000}\}$



Spectral Analysis



amplitude spectrum $$y_l = \sqrt{(y_l^c)^2 + (y_l^s)^2}$$

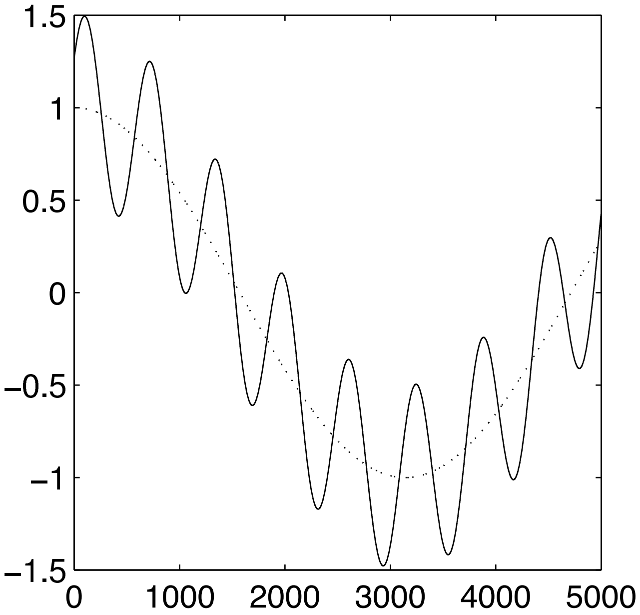
phase spectrum $$p_l = \arctan \frac{y_l^s}{y_l^c}$$

main periodic components of periodic data

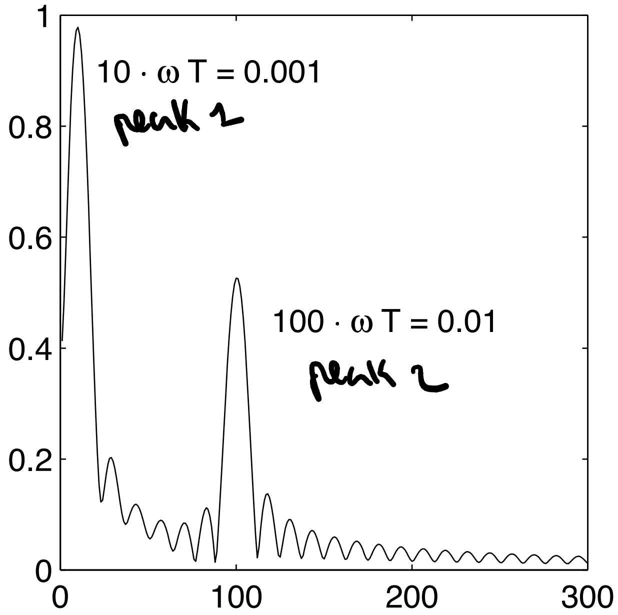
Example Amplitude and Phase Spectrum

$X = \{(i, \cos(0.001 \cdot i) + 0.5 \cdot \cos(0.01 \cdot i - 1)) \mid i \in \{1, \ldots, 5000\}\}$
$m = 300$ spectral values, $\omega T = 10^{-4}$

original data
$X = \{x_1, \ldots, x_{5000}\}$

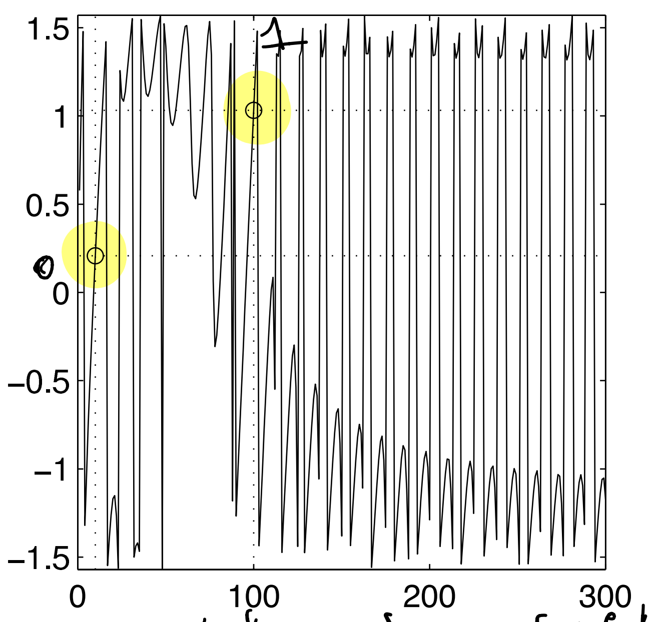


amplitude spectrum
$Y = \{y_1, \ldots, y_{300}\}$



frequency

phase spectrum
$P = \{p_1, \ldots, p_{300}\}$



relative phase shift

Prof. Dr. Thomas A. Runkler