Principal Component Analysis for Dimensionality Reduction of Sensor Data

Prof. Florian Römer
Ilmenau University of Technology / Fraunhofer Institute for Nondestructive Testing IZFP
florian.roemer@tu-ilmenau.de

Motivation Dimensionality Reduction

- Dimensionality in (SH) monitoring applications
  - **# of sensors**: depends, can be dozens, hundreds or more
  - **# of samples in time**: sampling rate x total sampling duration

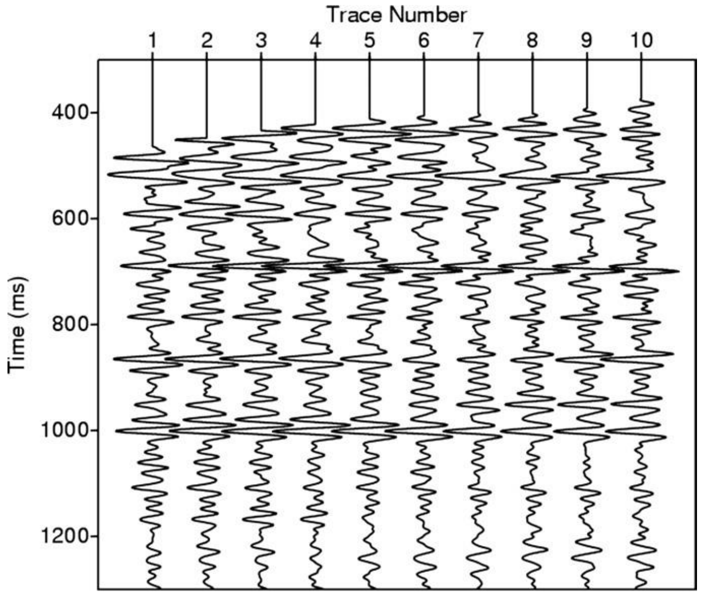
Depends:
- (mHz…Hz), e.g., temperature
- (kHz…MHz), e.g., ultrasonic signal

Depends, often in bursts:
- (ms…s) for single bursts
- (hours, days, months) for continuous monitoring!

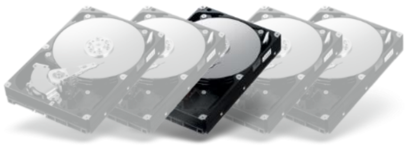
Monitoring systems quickly generate GB, TB, and more…

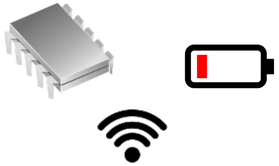
Motivation Dimensionality Reduction

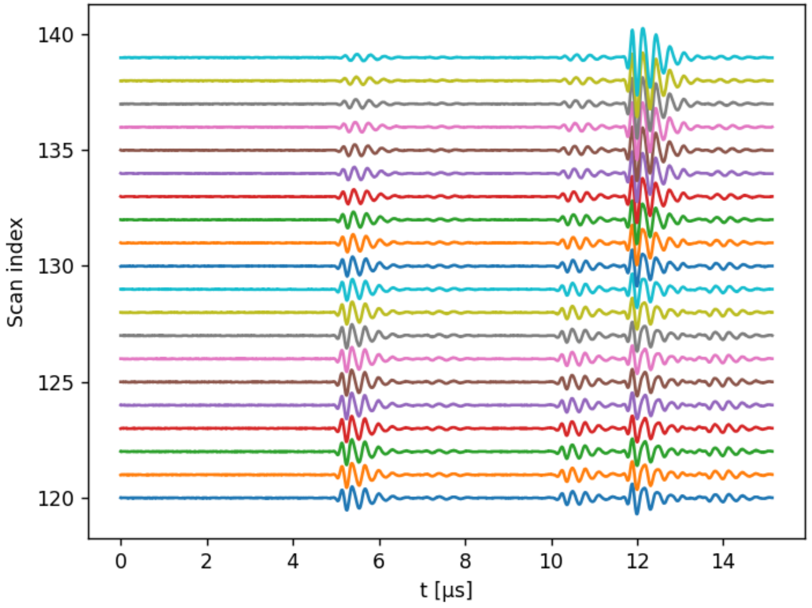
- At the same time, sensor data is often **highly redundant**
  - **Space**: adjacent scans similar (dense sampling of smooth wavefronts)
  - **Fast time**: resonant phenomena do not occupy full bandwidth, also oversampling not uncommon (e.g., for better identification of peaks). Isolated peaks lead to sparsity in time.
  - **Slow time**: in monitoring applications, the nominal signal stays constant for a long time until there is a relevant event
- Conversely, “non-redundant data” would fill R^N cube isotropically



- What benefits do we expect?
  - Reduction of **total data amount**: storing/maintaining large data is costly and inefficient
  - **Energy consumption**, in particular for autonomous sensor nodes
    - Measuring and storing data
    - Transmitting data!



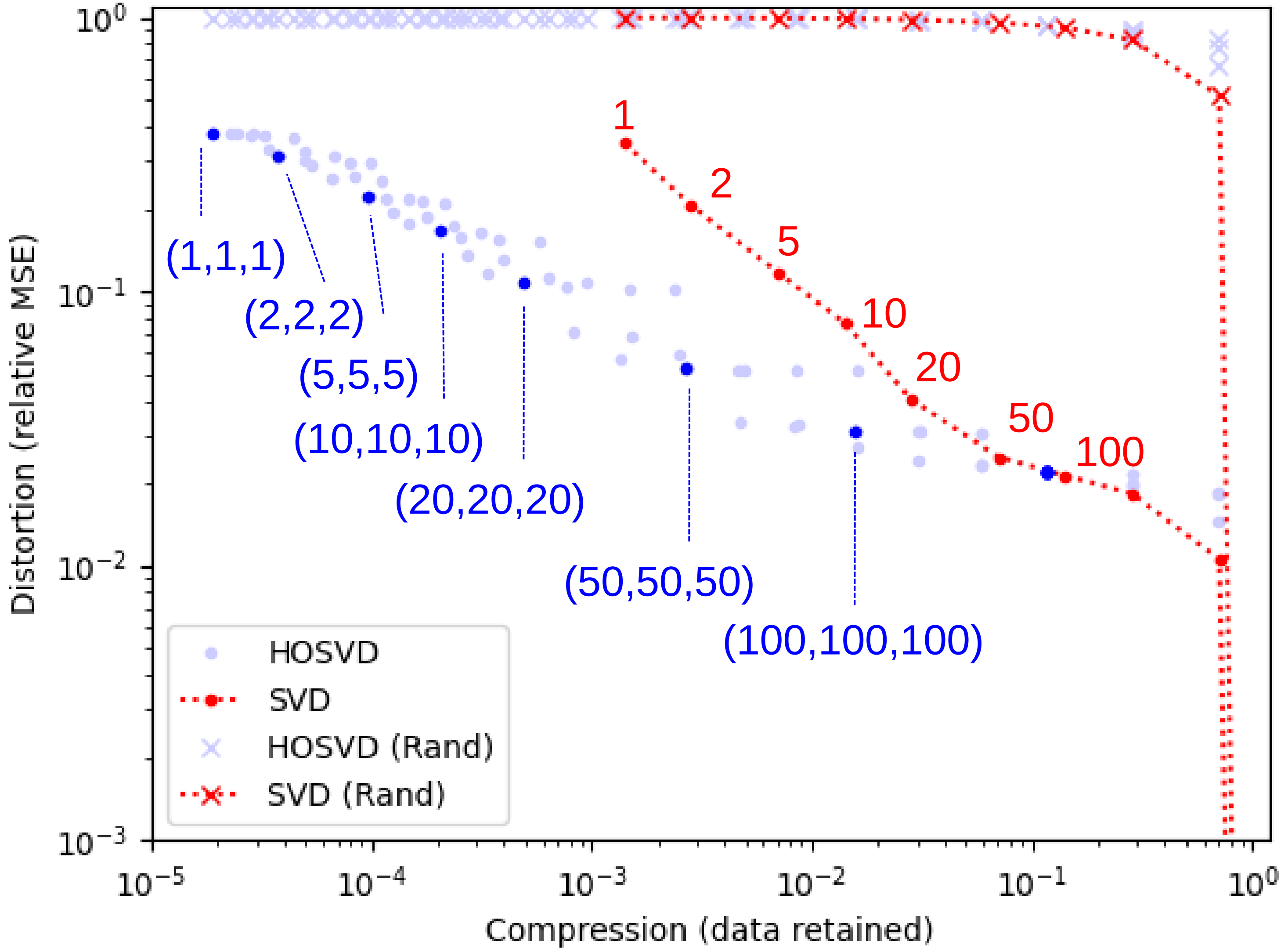




Learning Goals

- Understanding PCA
  - Low-dimensional subspace?
  - Singular vectors (data) vs. eigenvectors (correlation)?
  - Intuition: singular values / eigenvalues (scree plots)?
- Practical example
  - PCA synthetic data, ellipse plot
  - PCA Ultrasonic B-scan, reconstruction fidelity
- If time permits: Extensions
  - Kernel PCA
  - Multilinear PCA

3D PCA on a 3D Ultrasonic scan: PCA (SVD) vs HOPCA (HOSVD)



Data size SVD
1212 x 100965

Data size HOSVD
1212 x 265 x 381

Key Take-Aways

- Understanding PCA
  - Projects data onto the subspace that contains the largest amount of variance of the data.
  - Obtained from singular vectors of the data or eigenvectors the correlation matrix (Gramian)
  - Singular values allow us to directly judge amount of compressibility and select PCA rank
- Practical example
  - White noise is not compressible: principal components similar in strength, none of them negligible. If we correlate the noise, compressibility grows. PCA shows dominant „directions“ of the data
  - Ultrasonic B-Scan can be well compressed with a handful of components. Just the MSE is not enough, look at data fidelity.
- Extensions
  - Kernel PCA: bumping the data to higher dimenions can let us exploit nonlinear redundancies. In general better suited for classification than compression
  - Multilinear PCA: multidimensional correlations can be exploited more efficiently.