Barbara Hammer, Philip Kenneweg Universität Bielefeld

Modern Data Analysis (WiSe 2023/2024)
4. Sheet

Start: Wednesday, 22.11.2023.
End: The worksheets should be solved using Python, in groups of 2-3 people and will be presented in the Tutorials.
Discussion: Monday, 04.12.2023 in den Tutorien.

Information

The worksheets and necessary toolboxes will be made available in the Lernraum "392246 Modern Data Analysis (V)". Worksheets will usually be released every two weeks on Wednesday, and discussed during the exercises on Monday two weeks later. In order to successfully finish the course, 60% of the available points have to be obtained and each participant has to present his/her results at least once. The Monday in between the release and discussion of the sheet will be used to discuss the implementation of the various algorithms presented in the lecture.

Exercise 1: Compressed Sensing

(*10 Points*)

For the following exercise, you may use the sklearn OMP implementation.

(a) *(7 Pts.)* Load the file signal.mat. We have provided a signal $\vec{x} \in \mathbb{R}^{100}$. 10 associated sinusoidal base primitives are provided in $D \in \mathbb{R}^{100\times 10}$. $\vec{x}$ can be represented sparsely by $\vec{x} = D\vec{a}$ where $|\vec{a}|_0 = 3$.

First, use Orthogonal Matching Pursuit (OMP) to find the matching activations $\vec{a}$ and reconstruct $\vec{x_r} = D\vec{a} \approx \vec{x}$. The error $\sum_i |\vec{x}_i - \vec{x}_i^r|$ should be extremely small. Then, use the dictionary $D_2 = I \in \mathbb{R}^{100\times 100}$ to reconstruct $\vec{x}$. How does the error behave with respect to $|\vec{a}|_0$ and why does this behaviour differ from the behaviour of the first basis D?

As a next step, use Compressed Sensing to reconstruct the signal from very few measurements. Proceed as follows: Create an orthonormal basis $\Phi_{ext} \in \mathbb{R}^{100\times 100}$ of random measurement primitives by initializing uniformly distributed random vectors and orthonormalizing them (e.g. using scipy.linalg.orth). For a given number of measurements m, the "reduced" measurement basis $\Phi \in \mathbb{R}^{100\times m}$ consists of m columns that are drawn randomly from Φ_{ext}. The representation of the signal in this measurement basis is given by $\vec{y} = \Phi^T\vec{x}$. Use $\vec{y}$, Φ and D to compute the reconstruction $\vec{x_r}$ of the signal and compare the result with the original signal $\vec{x}$. Remember from the lecture: for sparse recovery we want to solve the problem:

find a where $|\vec{y} - \Phi^t Da|_2^2 \leq \epsilon$ and $|\vec{a}|_0$ is minimum.

using OMP and use these coeficients $\vec{a}$ to reconstruct $\vec{x_r} = D\vec{a}$

Repeat the process for several values of $m \leq 10$ and several random measurement bases. How many measurements are necessary to reconstruct $\vec{x}$ on average? Visualize your results throughout.

(b) *(3 Pts.)* Load the file image.mat. The vector $y \in \mathbb{R}^{500}$ contains the "measurements" of an unknown signal with regard to the random measurement basis $\Phi \in \mathbb{R}^{1024\times 500}$. The original signal was a gray value image with 1024 pixels [32 x 32], represented by a vector formed by concatenating the columns. Since the image consists of few strokes, a simple basis for sparsely representing the signal is the standard basis $D = I \in \mathbb{R}^{1024\times1024}$. Reconstruct the original picture using Compressed Sensing and display it. What does it show? How sparse is the signal in the chosen dictionary? Compute the reconstructed $\vec{x_r}$ using the pseudinverse e.g. $\vec{x_r} = \Phi^{t-1}\vec{y}$ and compare the results.