Exercises
Computational Intelligence Lab
SS 2019

Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 10, May 9-10, 2019
(Sparse Coding and Wavelets)

Problem 1 (Signal denoising):

1) Note that, since with $\mathcal{U}$ we can represent every element of $\mathbb{R}^D$, the noise $\boldsymbol{\epsilon}$ can be also expressed using dictionary elements. Hence, we have

$$\mathbf{x} = \sum_{k=1}^{D} \tilde{z}_k \mathbf{u}_k,$$

where the $\tilde{z}_k$ encode the noisy signal. Our estimate using $\hat{K}$ atoms and a permutation σ is

$$\hat{\mathbf{x}}_\sigma = \sum_{k=1}^{\hat{K}} \hat{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)}$$

By direct computation

$$\begin{aligned}
\left\|\mathbf{x} - \hat{\mathbf{x}}_\sigma\right\|_2^2 &= \langle \mathbf{x} - \hat{\mathbf{x}}_\sigma, \mathbf{x} - \hat{\mathbf{x}}_\sigma \rangle = \\
&= \left\langle \sum_{k=1}^{D} \tilde{z}_k \mathbf{u}_k - \sum_{k=1}^{\hat{K}} \hat{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)}, \sum_{k=1}^{D} \tilde{z}_k \mathbf{u}_k - \sum_{k=1}^{\hat{K}} \hat{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} \right\rangle = \\
&= \left\langle \sum_{k=1}^{\hat{K}} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} + \sum_{k=\hat{K}+1}^{D} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} - \sum_{k=1}^{\hat{K}} \hat{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)}, \sum_{k=1}^{\hat{K}} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} + \sum_{k=\hat{K}+1}^{D} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} - \sum_{k=1}^{\hat{K}} \hat{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} \right\rangle \\
&= \left\langle \sum_{k=1}^{\hat{K}} (\tilde{z}_{\sigma(k)} - \hat{z}_{\sigma(k)}) \mathbf{u}_{\sigma(k)} + \sum_{k=\hat{K}+1}^{D} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)}, \sum_{k=1}^{\hat{K}} (\tilde{z}_{\sigma(k)} - \hat{z}_{\sigma(k)}) \mathbf{u}_{\sigma(k)} + \sum_{k=\hat{K}+1}^{D} \tilde{z}_{\sigma(k)} \mathbf{u}_{\sigma(k)} \right\rangle
\end{aligned}$$

Next, using linearity of the inner product and the orthogonality of the atoms, one quickly gets

$$\left\|\mathbf{x} - \hat{\mathbf{x}}_\sigma\right\|_2^2 = \sum_{k=\hat{K}+1}^{D} (\tilde{z}_{\sigma(k)})^2 + \sum_{k=1}^{\hat{K}} \left(\tilde{z}_{\sigma(k)} - \hat{z}_{\sigma(k)}\right)^2.$$

Thus, clearly, one needs to choose $\tilde{z}_{\sigma(k)} = \hat{z}_{\sigma(k)}$ for $k = 1, \ldots, \hat{K}$. The cost is then simply $\sum_{k=\hat{K}+1}^{D} (\tilde{z}_{\sigma(k)})^2$. Moreover,

$$\hat{\sigma} = \operatorname*{argmin}_{\sigma} \left\{ \sum_{k=\hat{K}+1}^{K} (\tilde{z}_{\sigma(k)})^2 \right\} \tag{SOL}$$

It is obvious that the above expression is minimized when the permutation sorts the *noisy* coefficients in order of decreasing magnitude.

2) Let us compute the power (i.e. the magnitude squared) of each coefficient in the spectrum of $\mathbf{x}$:

$$\left(\tilde{z}_{\sigma(k)}\right)^2 = \left(\mathbf{u}_{\sigma(k)}^T \mathbf{x}\right)^2 = \left(\mathbf{u}_{\sigma(k)}^T \left(\sum_{k=1}^{\tilde{K}} z_{\sigma^*(k)} \mathbf{u}_{\sigma^*(k)} + \boldsymbol{\epsilon} \right)\right)^2 = \begin{cases} \left(z_{\sigma^*(k)}\right)^2 + \boldsymbol{\epsilon}^T \boldsymbol{\Pi}_{\mathbf{u}_{\sigma(k)}} \boldsymbol{\epsilon} & \text{if } \sigma(k) = \sigma^*(k) \\ \boldsymbol{\epsilon}^T \boldsymbol{\Pi}_{\mathbf{u}_{\sigma(k)}} \boldsymbol{\epsilon} & \text{otherwise} \end{cases}.$$

Where $\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}} := \mathbf{u}_{\sigma(k)}\mathbf{u}_{\sigma(k)}^T$ is the projection matrix onto the linear space spanned by $\mathbf{u}_{\sigma(k)}$. Next, we want to compute the expected value of each coefficient. First, notice that, since $\boldsymbol{\epsilon}^T\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\boldsymbol{\epsilon}$ is a number,

$$\mathbb{E}\left[\boldsymbol{\epsilon}^T\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\boldsymbol{\epsilon}\right] = \mathbb{E}\left[\text{trace}\left(\boldsymbol{\epsilon}^T\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\boldsymbol{\epsilon}\right)\right].$$

By the ciclicity of the trace, (trace$(AB) =$ trace(BA) for any matrices A and B for which AB is a square matrix) and linearity of expectation and of the trace

$$\mathbb{E}\left[\text{trace}\left(\boldsymbol{\epsilon}^T\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\boldsymbol{\epsilon}\right)\right] = \mathbb{E}\left[\text{trace}\left(\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right)\right] = \text{trace}\left(\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right]\right).$$

Finally, notice that $\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right]$ is the covariance matrix of the noise, which we assume white Gaussian. Therefore $\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right] = \gamma^2\mathbf{I}$ for some small γ^2. Hence we can simplify:

$$\text{trace}\left(\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right]\right) = \text{trace}\left(\gamma^2\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\right) = \gamma^2\text{trace}\left(\mathbf{\Pi}_{\mathbf{u}_{\sigma(k)}}\right) = \gamma^2,$$

where in the last equality we used the fact that projection matrices on one dimensional spaces have unit trace; indeed, for any $\mathbf{v}$ with unit norm

$$\text{trace}\left(\mathbf{\Pi}_{\mathbf{v}}\right) = \text{trace}\left(\mathbf{v}\mathbf{v}^T\right) = \text{trace}\left(\mathbf{v}^T\mathbf{v}\right) = \mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|^2 = 1.$$

All in all, we get

$$\mathbb{E}\left[\tilde{z}_{\sigma(k)}^2\right] = \begin{cases} \left(z_{\sigma^*(k)}\right)^2 + \gamma^2 & \text{if } \sigma(k) = \sigma^*(k) \\ \gamma^2 & \text{otherwise} \end{cases}.$$

Thus, by comparing the last equation with (SOL), we conclude that if we know the correct $\tilde{K}$ (that is $\hat{K} = \tilde{K}$), in expectation the solution to (P) is correct: $\mathbb{E}[\hat{\sigma}] = \sigma^*$ (up to a permutation of the first $\tilde{K}$ indices).

Problem 2 (1D signal compression and Haar wavelets):

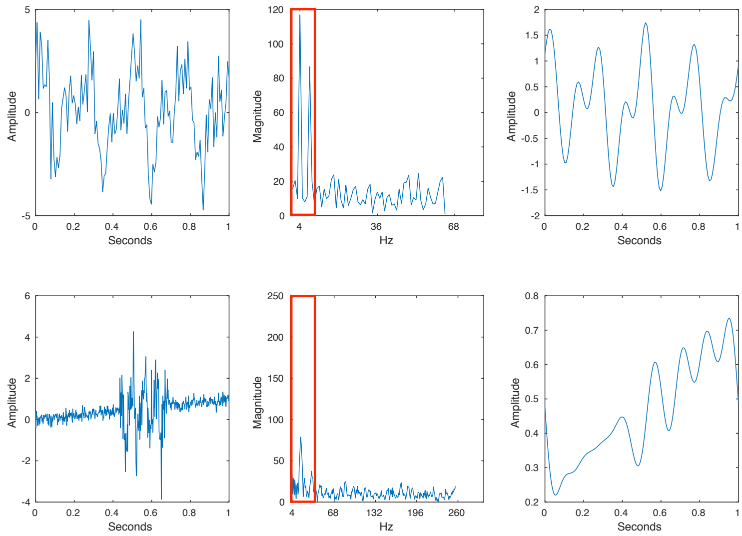
Please the solution at `ex2-sol.ipynb` from

github.com/dalab/lecture_cil_public/tree/master/exercises/ex10/ex2-sol.ipynb.

Problem 3 (Choice of dictionary is crucial):

The figure above shows two different 1-D signals (left column) with their corresponding spectrum obtained using the FFT (middle column). In the right column, we show a signal obtained by discarding part of the frequencies in the spectrum.

- (1) Write down the formula to obtain the spectrum in the middle column of the previous figure, in terms of linear transformation or change of basis (assuming a given basis $\mathbf{U}$) applied to the original signal $\mathbf{x}$.
  Solution:
  $$\mathbf{z} = \mathbf{U}^T\mathbf{x}$$
  (2) Write down the inverse formula to obtain the reconstructed signal in the right column in terms of linear transformation (change of basis) applied to the filtered spectrum $\hat{\mathbf{z}}$.
  Solution:
  $$\hat{\mathbf{x}} = \mathbf{U}\hat{\mathbf{z}}$$
- What part of the signal would you discard to obtain the reconstructed signal? Draw a rectangle on each spectrum in the middle column where everything inside the rectangle is kept for the reconstruction.
  Solution:



- **False** The Fourier transform is a better choice than Wavelet for the first signal because the frequency components in the signal are global.

- **True** The Wavelet transform is a better choice than Fourier for the second signal because the signal has localized frequency components.

- The first peaks in the spectrum correspond to the low-frequency components.

Problem 4 (Image compression):

Please find the solution in the iPython notebook `ex4-sol.ipynb` from

github.com/dalab/lecture_cil_public/tree/master/exercises/ex10/ex4-sol.ipynb,

and answer the questions.