MCB 131: Problem Set 2

Due: Monday, 6 March 2017 2:30PM

Problem 1. Infomax for natural images

- In class we have derived the Infomax cost function,

$$E = -\sum_k \left\{ \frac{1}{2} \ln[x_k(c_k+1)+1] - \frac{1}{2} \ln[x_k+1] - \frac{1}{2}\rho[x_k(c_k+1)+1-\sigma^2] \right\} \tag{1}$$

  where the sum is over all M Fourier wave vectors k. (Note that k is a 2D vector and for simplicity of notations, we use k instead of $\vec{k}$, similarly for r below.) As we did in the class, we will mainly work with the scaled parameters as defined below.

  (i) c_k is the scaled power spectrum of natural images, C_k/ϵ, where ϵ is the input noise variance,

  (ii) x_k is $|w_k|^2$ where w_k is the Fourier transform of the (scaled) localized receptive field, $w(r)$ (the unscaled version is $w(r)/\sqrt{\epsilon}$), and

  (iii) the output noise variance is kept fixed at 1. $\sigma^2 > 1$ is the value of the total output variance (for the unscaled variables) divided by M and ρ is the corresponding Lagrange multiplier for constraining this total output.

  - A. By evaluating $\partial E/\partial x_k = 0$ derive the following solution for x_k

$$x_k = \frac{1}{2(c_k+1)} \left[-(c_k+2) + \sqrt{(c_k+2)^2 + 4\left(\frac{c_k}{\rho} - c_k - 1\right)} \right]$$

  - B. Use the above equation to show that (e.g. by Taylor expansion) in the low noise limit, $\epsilon \to 0$, i.e., $c_k \to \infty$,

$$x_k \approx \frac{1}{c_k}\left[\frac{1}{\rho} - 1\right] > 0 \tag{2}$$

    while in the high noise limit $\epsilon \to \infty$, i.e., $c_k \to 0$,

$$x_k \approx \left[\sqrt{\frac{c_k}{\rho}} - 1\right]. \tag{3}$$

    Hint: In deriving these limits, you have to be mindful of how ρ behaves in those limits. It can be shown that in the low noise limit ρ remains of order 1, while in the high noise limit, $\rho \propto 1/\epsilon$, so that $c_k/\rho = O(1)$.

  - C. Evaluate $|w_k|$ vs. k for natural images and different levels of noise.

    Details: Assume the signal is arranged in a square lattice with unit distance between nearest neighbors. (Choose the linear size $L = 101$ or something of this order). For $c_k = C_k/\epsilon$ the following form can be used,

$$C_k = \frac{1}{(k^2 + k_{min}^2)^{0.9}} \tag{4}$$

    Note that we have added a “low frequency cut-off”, $k_{min} = 2\pi/L$, to prevent C_k from diverging at $k = 0$.

(i) Recall that $|w_k| = (x_k)^{1/2}$ and use the expression derived in question A to evaluate $|w_k|$ vs. $|k|$ for the following noise levels: $\epsilon = [0.01,\ 1,\ 20 : 40 : 380]$. You can plot for $k = (k_x, 0)$, for simplicity, or better by making a 2D surface plot. Use the values of ρ given in the mat file **Rho_vs_ep.mat** (or text file **Rho_vs_ep.txt**) which gives the values of ρ for the relevant values of ϵ.

(ii) Plot the (scaled) receptive field $w(r)$ in real space for low and high noise, for example: $\epsilon = 0.01$ and $\epsilon = 100$ by Inverse Fourier Transform of w_k (assume the phases are zero). You can plot 2D surfaces or for one dimensional cross section e.g., $r = (r_x, 0)$.

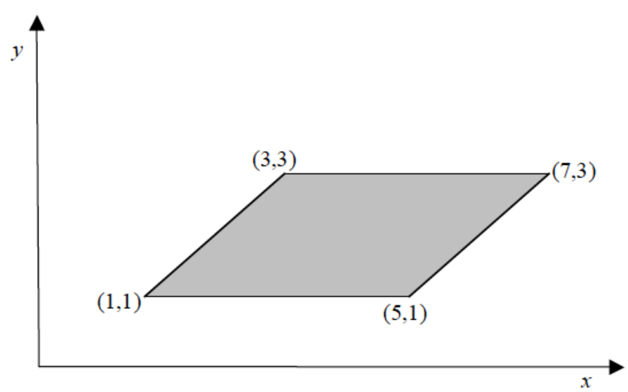
(iii) The provided ρ values are calculated at $\sigma^2 = 4$. Check that your solution indeed obeys (up to some reasonable precision) the total output variance constraint. [Hint: can you think of a shortcut to derive the equation of this varaince constraint using Eq.(1) ?]

Problem 2.

Consider an image whose pixels form a one dimensional lattice of points with periodic boundary conditions. The intensities of the pixels are given by the vector $x = (x_1, x_2, \ldots, x_N)^T$ with zero mean and covariance matrix with the following structure: C_{ij} =1 for all $i = j$. All off-diagonal elements of C are zero except for $|i - j| = 1$ (where $|i - j|$ is mod N) in which case $C_{ij} = a$, $a < 0.5$.

- (a) Using the fact that C as defined above is a circulant matrix, give an analytical expression for all the eigenvalues and eigenvectors of the matrix C. What is the degeneracy of the eigenvalues (i.e., how many of them are equal to each other)?
- (b) Plot the eigenvalues as a function of the Fourier number k, for $N = 100$ and several positive and negative values of a.
- (c) Verify your results by diagonalizing the matrix numerically with the Matlab command *eig*.
- (d) Explain why a is restricted to be smaller than 0.5.

Problem 3.



Given an ensemble of two-dimensional input vectors $r = (x, y)$ that take on values distributed uniformly inside the above parallelogram, inputing feedforward network of two linear output neurons with weight vectors W^1 and W^2, respectively:

- (a) Find the principal components for the distribution, PC1 and PC2. How much variance is accounted for by each PC? Note: First, shift the parallelogram so that it is centered around the mean vector, so that the new r has zero mean.

- (b) Write the general form of W^1 and W^2 that whiten the input. (Hint: Note that the general form can be written as $W_{PCA}U$ where W_{PCA} denotes what we called in class the PCA solution and U is a rotation matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ denoting a counterclockwise rotation by θ.) Draw a couple of examples of the resultant vectors W^1 and W^2. Directly compute the covariance of the output neurons for all the solutions and show its white.

- (c) Statistical Dependence: Prove that in general, the output neurons, $y^1$and y^2 are statistically dependent by showing that in general, given the value of y^1, impose some restriction on the range of values that y^2 can have. Suggest a high order correlation that will capture this dependency. However, there is one value of θ for which knowing y^1 does not tell you anything about y^2. Determine the value of θ and the associated two vectors W^1 and W^2. By Drawing them you should find geometrically what is special about this pair of filters. [It might be amusing and helpful to write down a simple script that generates the pairs $W^1$and W^2 from randomly generated points uniformly distributed in the parallelogram].

Problem 4. Independent component analysis and blind source separation

Section A: Skewness and kurtosis of a mixed random variable

Consider a single random variable

$$x = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} a_i \, s_i$$

where s_i are i.i.d. (independent, identically distributed) random sources. We denote the lowest order moments of s_i by:

$$\mu_1 = \langle s_i \rangle = 0,\, \mu_2 = \langle s_i^2 \rangle,\, \mu_3 = \langle s_i^3 \rangle,\, \mu_4 = \langle s_i^4 \rangle$$

Assume throughout Section A that all components of the weight vector, a_i, are equal to $+1$ or -1.

a) Compute the mean, variance, skewness, and kurtosis of x.

b) How do these quantities depend on N and on the particular choice of a? Explain the significance of your results for ICA.

Reminder: The skewness and kurtosis of the random variable x are defined as $\gamma = \frac{\langle x^3 \rangle}{\langle x^2 \rangle^{3/2}}$ and $k = \frac{\langle x^4 \rangle}{\langle x^2 \rangle^2} - 3$, respectively.

Section B: Numerical demixing algorithm

In this exercise you will implement a numerical method for blind source separation by maximizing kurtosis. Download "mixed_images.mat" from the course website. This matfile contains three images that were generated by linearly mixing three (unknown) source images. You will find the three source images using an ICA algorithm. For Section B, please turn in your Matlab code and plots of the requested figures. Instructions:

1. Load the .mat file in Matlab and plot the mixed images (use the "gray" colormap). Examine the pixel statistics of each image (mean, variance, skewness and kurtosis).

2. Frame the problem as an ICA problem. The input vectors are 3-D, consisting of the three image intensity values corresponding to a pixel location. You have to find a 3×3 matrix W such that $\mathbf{u} = W\mathbf{x}$ will give the corresponding intensity values of the three source images.

3. As a first step - spherize (i.e, subtract the mean and whiten) $\mathbf{x}$.

4. A simple algorithm for finding the demixing matrix W is to sequentially find each of the three (normalized) demixing vectors $(\mathbf{w_1}, \mathbf{w_2}, \mathbf{w_3})$ that comprise W. Choose some initial value for the first 3-D filter vector, $\mathbf{w_1}$. Update $\mathbf{w_1}$ by adding it to a term proportional to the gradient of the kurtosis of $\mathbf{w_1} \cdot \mathbf{x}$ and then normalize $\mathbf{w_1}$. After repeated iteration, this should converge to one of the source images, which should have greater kurtosis than any of the mixed images. Plot the source image. Plot the kurtosis as a function of the iteration number.

5. Repeat this procedure to find $\mathbf{w_2}$ by maximizing the kurtosis in the space perpendicular to $\mathbf{w_1}$. You can achieve this by first updating $\mathbf{w_2}$ with the gradient of the kurtosis and then subtracting the component lying in the direction of $\mathbf{w_1}$. This should converge to the second source image. Plot the second source image. Plot the kurtosis as a function of the iteration number.

6. Finally, obtain $\mathbf{w_3}$ by maximizing the kurtosis and subtracting out the projection on the 2-D space spanned by $\mathbf{w_1}$ and $\mathbf{w_2}$. Plot the third source image. Plot the kurtosis as a function of the iteration number.

EXTRA CREDIT:

You may submit at most one of the following problems to be graded as extra credit.

1 Compressible Signals

Consider a rank ordered signal $x \in R^N$ $(x_i \geq x_{i+1}\ \forall i)$.

- Show that the vector x_K,

$$(x_K)_i = \begin{cases} x_i & i \leq K \\ 0 & i > K \end{cases}$$

  is the $K-$sparse vector with the minimal L_2 distance from x. (i.e $x_K = \arg\min_{y \in K-\text{sparse}} \|y - x\|_2)$. $x_K$is the best $K-$sparse approximate of x .

- One of the basic results of CS theory is that the error of L_1reconstruction of any signal is bounded by

$$||\hat{x} - x||_2 \leq C_1 \frac{||x - x_K||_1}{\sqrt{K}} \tag{5}$$

where $\hat{x}$is the reconstructed signal and C_1 is some positive K independent constant.

- Consider the case where for some constant R: $\forall i,\ |x_i| \leq \frac{R}{i^p}$, with $p \geq 1$.
  Show that in the limit $N, K \gg 1$, $\frac{K}{N} \sim O(1)$, and $p \geq 1$, the following relation holds

$$\frac{\|x - x_K\|_1}{\sqrt{K}} \simeq C \|x - x_K\|_2 \ ,$$

  where C is some $O(1)$ constant.

- Show that the above relation does not hold if $p < 1$.

- Discuss why signals with $p \geq 1$ are called *compressible signals.*

2 IRLS Algorithm and Compressed Sensing Simulations

The IRLS algorithm[1] minimizes the L_1 norm of a vector subject to linear equality constraint by iteratively solving weighted least square optimization problems. i.e. solving

$$\min_x \sum_{i=1}^{N} \omega_i^{(n)} x_i^2 \text{ s.t. } Ax = y$$

with the weights of the n iteration given by $\omega_i^{(n)} = \frac{1}{\left|x_i^{(n-1)}\right|}$ were $x_i^{(n-1)}$ are the solution of the optimization problem in the previous iteration. By writing the Lagrangian of the optimization problem and minimizing it, one finds that

the solution in each iteration is given by

$$x^{(n)} = Q_{(n)}^{-1} A^T \left(A Q_{(n)}^{-1} A^T \right)^{-1} y$$

with $Q_{(n)} = \operatorname{diag}\left(\omega^{(n)}\right)$.

.

In this question you will find by simulation the minimal number of measurements required to reconstruct a sparse signal by using L_1 minimization.

Consider a $K-$sparse signal, s, of size N with K non zero element drawn randomly by first choosing the non zero elements uniformly and then drawing the value of the non zero elements from a standard normal distribution.

We will try and reconstruct the signal from M measurements given by

$$y = As$$

with the matrix $A_{M\times N}$ a random Gaussian matrix with an appropriate choice of mean and variance[1], using the IRLS algorithm .

- Download the file IRLS.m from the course website. The file contains a MATLAB code that implements the IRLS algorithm for a given matrix A and a given measurement vector y.
- For a given sparsity $f = \frac{K}{N}$ calculate the probability for a perfect reconstruction of the signal as a function of the fraction of measurements $\alpha = \frac{M}{N}$. How does the result depend on N? (Use $f = 0.15$ and $N = 100,\ 250,\ 500$)
- Measure α_c, the critical value of α above which perfect reconstruction is guaranteed with high probability for large N, as a function of the sparsity f. Compare your results to Figure 3 in the CS review by Ganguli and Sompolinsky [2].

Recommended reading for extra credit (available on the course website)

1. Statistical mechanics of compressed sensing. S. Ganguli and H. Sompolinsky, Phys. Rev. Lett. (2010) 104:188701.
2. Iteratively reweighted algorithms for compressive sensing. R. Chartrand, and W. Yin, IEEE (ICASSP). (2008) pp 3869–3872.