1 Theory

1. Show that
$$\nabla_{\mathbf{v}}(\mathbf{v},\mathbf{v}) = 2\mathbf{v}$$
and that if C is a symmetric matrix, then
$$\nabla_{\mathbf{v}}(\mathbf{v},C\mathbf{v}) = 2C\mathbf{v}.$$

2. Show that
$$(\phi^{(1)}, C\phi^{(2)}) = (C\phi^{(1)}, \phi^{(2)}).$$
Assume C is symmetric.

3. Given the data matrix
$$X = \begin{pmatrix} -2 & -1 & 1 \\ 0 & -1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 1 \end{pmatrix},$$
compute the eigenvalues and eigenvectors of XX^T and X^TX. For $\mathbf{u}^{(1)}$, confirm the statement
$$\mathbf{u}^{(j)} = \frac{1}{\sigma_j}\sum_{k=1}^{P} v_k^{(j)}\mathbf{x}^{(k)},$$
where $j = 1,\ldots,rankX$.

4. It was shown that the expansion coefficients may be computed using formula
$$A = \Sigma V^T,$$
providing an alternative to the direct computation via
$$A = U^T X.$$
Compute the number of add/multiplies required to compute A via both formulas, assuming that A is of size $N \times P$. Which way is computationally cheaper in general? Why?

2 Computing

1. The object of this programming assignment is to write a code to apply the *snapshot* method to a collection of P high-resolution image files. Your program should compute (and order) the eigenpictures. It should also have a subroutine to determine the projection of a given picture onto the best D-element subspace (D is typically chosen empirically). Your program report should include the following information:

   (a) A display of the ensemble-average image.

   (b) A picture of a mean-subtracted image, for one of the images chosen at random from the ensemble.

   (c) A collection of eigenpictures (based on mean-subtracted data) for a broad range of eigenvalues. The eigenpictures must be mapped to integers on the interval $[0,255]$.

(d) Partial reconstructions of a selected image for various value of D. Include the reconstruction error $||\mathbf{x}-\mathbf{x}_D||$ in each case and confirm that you obtain perfect reconstruction when D is equal to the rank of the data matrix.

(e) A graph of $\lambda_i/\lambda_{\max}$ vs i, where λ_i is the i^{th} eigenvalue of the mean-subtracted, ensemble-averaged covariance matrix. How does this plot help you determining the best D value to use?

(f) Now, devise (describe) a classification algorithm that uses this idea of best basis to classify a probe (testing) data against a given gallery. (For your reference: this process is called *Principal Component Analysis*.) Why is this more efficient than classifying data points in their resolution dimension?

The data for this problem may be downloaded from the course website. The data file faces1.mat contains 109 images whose dimensions are 120×160. It is a single matrix, where each column has length $19,200$, which is 120×160. The format of the data is "uint8", which stands for unsigned integer, 8 bits. Before you use the data for KL, change it to "double" format.

2. Test your theory from 1(f) on the following data set: Digits.mat can be downloaded from the course website. It contains three variables: Gallery, Probe, and photo_size, where Gallery is a 1024×500 matrix with 50 digits of 0 in its first 50 columns, 50 digits of 9 in its last 50 columns, etc. The row dimension comes from the resolution of the images stored in photo_size. The variable Probe stores a set of novel digits from 0 to 9 that do not appear in the Gallery. Use *Principal Component Analysis* to classify the probe images against the gallery images. How well did the algorithm perform? Report and analyze your result.