HST.582J / 6.555J / 16.456J Biomedical Signal and Image Processing
Spring 2007

HST-582J/6.555J/16.456J Biomedical Signal and Image Processing Spring 2007

Laboratory Project 5
Blind Source Separation:
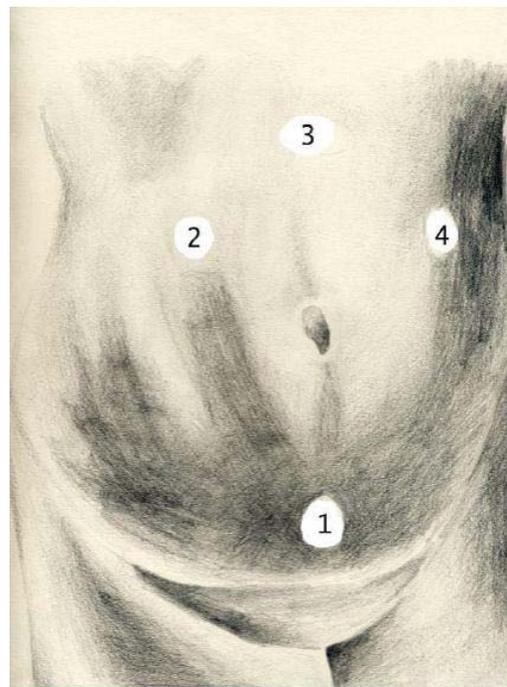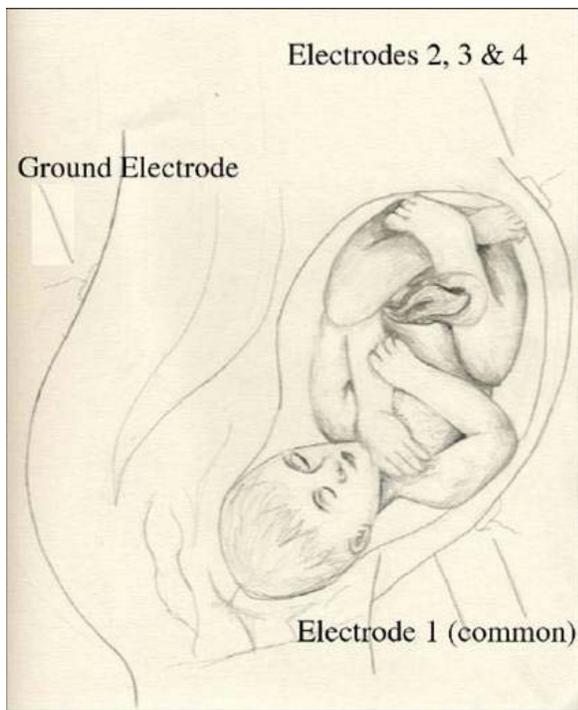### Fetal & Maternal ECG

DUE: 5/17/07

1   Introduction

Many health-related problems that occur during a human's lifespan (particularly in the first few months of life) can be reduced if related signs and symptoms are noticed during pregnancy. Since heart rate is one of the most important physiological indicators of health, accurately determining a fetus' heart rate is an important task. One important low-cost non-invasive method for deriving the HR is by recording and processing the electrocardiogram (ECG). If ECG sensors are placed on the mother's abdomen (usually in a *kite* configuration - see Fig. 1), then it is possible to record multiple differential signals between each pair of electrodes. In this example there are three possible *leads*[1] between each of the positions labeled 2, 3 and 4. An example of a fetal ECG (fECG) is given in figure 2 (left). Note the lower amplitude and simplified nature of the signal in comparison to the maternal ECG (mECG) in Fig. 2 (right).

ECGs are sensitive to surrounding contaminants such as measurement noise, electrode movement (due to shifts in the mother and fetus) and the mother's ECG itself. Fig. 3(i) illustrates a typical signal recorded from the configuration in figure 1. This is combination of the maternal ECG [Fig. 3(ii)], observation noise [Fig. 3(iii)], and fetal ECG [Fig. 3(iv)]. Note that, apart from the morphology differences, the fECG has a smaller amplitude (about one-third the height of the maternal QRS complex), which is nearly lost in the background noise. Furthermore, the fetus has a much higher heart rate than the mother. However, the spectral content of the mECG and fECG are similar. This makes filtering in the frequency domain extremely difficult.

Conventional methods for separating the mECG from the fECG involve a combination of bandpass filtering, maternal template matching, and signal subtraction. Unfortunately this leads to severe distortion of the fetal ECG and only the derived heart rate is useful. In this lab we will attempt to use Wiener filtering, Single Value Decomposition (SVD), and Independent Component Analysis (ICA) to separate the fECG from the noisy mixed signal.

---

[1]The term *lead* refers to the signal that constitutes the fluctuations in the potential difference between two electrodes. Confusingly, the wire attached to each electrode is sometimes referred to as a lead, but this is not what is meant in a clinical sense.

Courtesy of B. Campbell. Used with permission.

Figure 1: Standard electrode positions on the mother's abdomen for measuring three channels of fetal ECG (three over the fetus plus one common electrode just above the pubic bone and one ground electrode on the mother's lower back). Each observed ECG is a record of the mV fluctuations on the surface of the mother's abdomen between each pair of electrode positions (2, 3,and 4). Images ©B. Campbell, reproduced with permission.
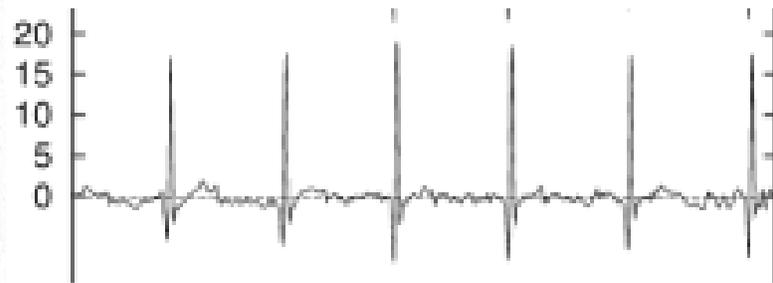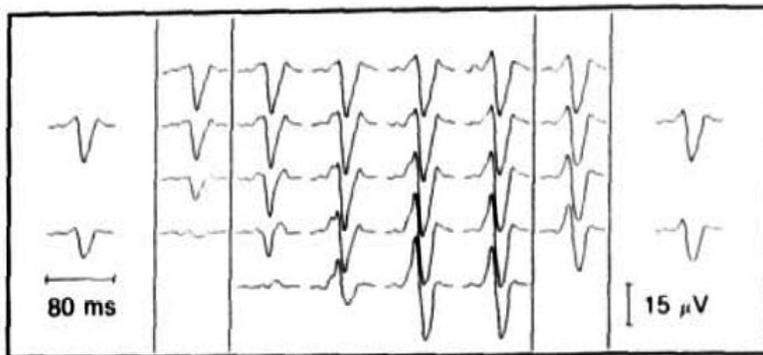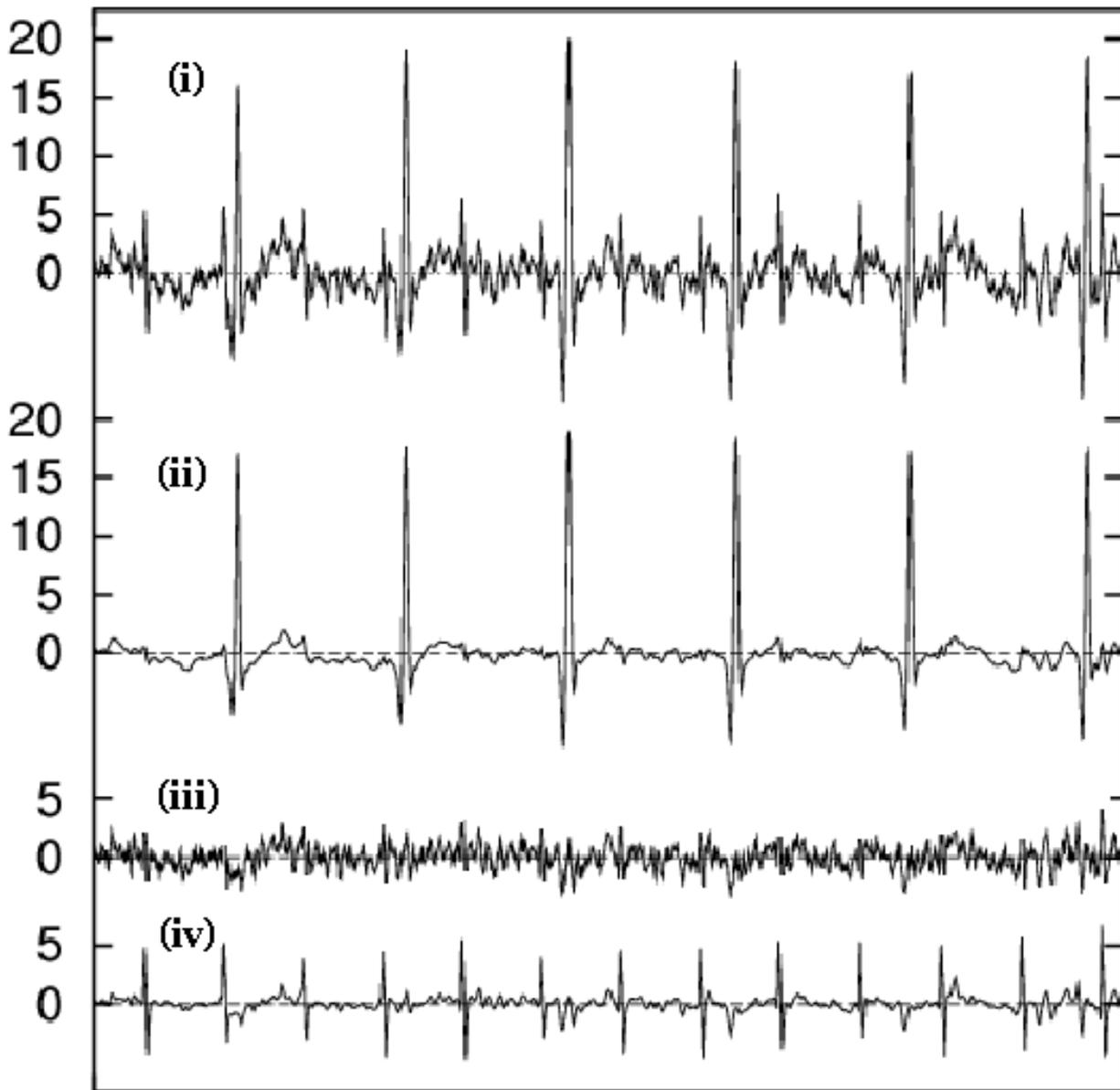

Figure 4(a), p. 152 from Oostendorp T., and A. vanOosterom. "The Fetal ECG throughout the Second Half of Gestation." *Clin Phys Physiol Meas* 10, no. 2 (1989): 147–160. Courtesy of IOP Publishing, Inc. Used with permission.

Richter M., T. Schreiber, and D. T. Kaplan D T. "Fetal EEG Extraction with Nonlinear Phase Space Projections." *IEEE Trans Bio Med Eng* 45 (January 1998): 133–137. Copyright © 1998 IEEE. Used with permission.

Figure 2: Typical fetal ECG trace (left) recorded at 26 weeks gestation and a typical maternal ECG (right). Note that the fECG morphology often changes significantly from the above trace after about 30 weeks. Adapted from Oostendorp *et al* [1] and [2].

2

Figure 3: From top: (i) Mixture of maternal and fetal ECG, (ii) maternal ECG only (iii) noise (iv) fetal ECG. The window is 5 seconds long for this trace and therefore the maternal heart rate is 72bpm and the fetal heart rate is 156bpm. Adapted from [2].

2 Laboratory exercises

In this laboratory, you will attempt to separate the components of a fetal ECG recording: the maternal waveform (like that in Fig. 3(ii)), observation noise (Fig. 3(iii)) and the fetal ECG itself (Fig. 3(i)). You will investigate three methods of signal separation: Wiener Filtering, Principal Component Analysis (using Singular Value Decomposition), and Independent Component Analysis.

The first two methods (Wiener Filtering and Principal Component Analysis) only use second order statistics. In contrast, Independent Component Analysis uses higher order statistics to perform signal/noise separation. In this lab you will see how the statistical distributions of the signals to be separated impacts the effectiveness of the three signal separation techniques.

In practical applications, a major obstacle to assessing how well signal separation techniques perform is our lack of knowledge regarding the true underlying signals. To overcome this problem we will use artificial ECG and noise signals so that we completely know our (uncontaminated) source signals and can therefore estimate how well our three techniques perform in separating out the source signals.

2.1 Preliminary Instructions and Observations

After logging on to your Athena workstation, type the following at a command prompt:[*]

```
add matlab
add 6.555
cd [my working directory]
matlab &
```

At the Matlab prompt type the following:

```
addpath /mit/6.555/data/bss/
addpath /mit/6.555/matlab/bss/
```

Begin by examining the maternal and fetal ECG signals as well as an example of measurement noise. The data is stored in **mecg1.dat**, **fecg1.dat**, and **noise1.dat** and can be read into Matlab using the **load** command. (Don't forget to include the *.dat* extension in the filename.) Each file consists of a vector of the components of a 'recorded' ECG signal (in mVs), sampled at a rate of 256 Hz.

Add the three signals to create a new signal that simulates what we might actually observe in a clinical environment.

[*] Note for OCW users: Athena is MIT's UNIX-based academic computing facility. The necessary Athena-based files and directories referenced herein have been provided in the supporting ZIP archives.

4

Question 1 *Include in your lab report separate time domain plots of the maternal, fetal, and noise signals, as well as the simulated 'observation'. What is the heart rate of the maternal and fetal ECGs? Identify, by inserting arrows on your plot, the fetal QRS complexes in the simulated observation.*

Question 2 *Calculate and plot the power spectrum of the maternal, fetal, and noise signals. (Use the Matlab function **pwelch**.) How do the fetal and maternal ECGs compare in the frequency domain? Describe the similarities and differences in one or two sentences.*

Question 3 *Calculate the first two moments (mean and variance) of the maternal, fetal, and noise signals. How do they compare? What does the variance tell you about the signals' power spectra?*

Question 4 *Use the Matlab function **hist** to plot an estimate of the PDF of each source signal and include these plots in your lab writeup. Calculate the 4th moment of the signal distributions, κ, using the Matlab function **kurtosis** and record these quantities. What does the value of κ tell you about the Gaussianity[2] of each signal? Relate this to your estimates of the PDFs (histograms). (Suggested length: 2 or 3 sentences).*

2.2 Wiener filtering of the Maternal ECG

Recall (from Chapter 12 of the lecture notes) that for the application of filtering out additive noise, the optimal (non-causal) Wiener filter is given by

$$H(f) = \frac{S_y(f)}{S_y(f) + S_d(f)} \tag{1}$$

where $S_y(f)$ is the power spectrum of the desired signal, $y[n]$, and $S_d(f)$ is the power spectrum of the noise component, $d[n]$.

Your task is to build a Wiener filter for this data and to apply it to the noisy observation you constructed in question 1. Copy the function template **wienerFilter.m** to your working directory, edit it to make it work correctly, and run it on the noisy observations. The function requires two inputs, a segment of the desired signal we wish to recover and the noisy observation. Note that for our application the fetal ECG is the desired signal and the additive 'noise' is the sum of the maternal and noise components; we are treating the maternal ECG as a noise component in this application!

Question 5 *Look at the filtered signal in both the time and frequency domain and include relevant plots in your report. Compare the desired fetal ECG signal to the filtered output. How well did your Wiener filter perform? (Suggested length: 1 or 2 sentences).*

---

[2]Note that MATLAB uses the convention that a Gaussian signal has a kurtosis of 3 (rather than zero).

Note: Be sure to save the output of your Wiener filter, as you will need to compare it with the results from other techniques later in this lab.

We have already seen that the underlying source signals are all zero mean with unit variance. However, differences in distribution of signal energy as a function of frequency allow us to separate the fetal ECG out (to some extent), provided that we know the approximate form of the power spectra and provided that they are different. Another method of filtering/separating signals that does not use any prior knowledge of the spectral components of the signals is Singular Value Decomposition (SVD), which we explore next.

2.3   Signal separation using SVD

SVD is a powerful, yet standard, second order technique for filtering data based upon the projection of the data onto orthogonal axes corresponding to vectors of maximal variance. Although this means we are projecting data onto a set of orthogonal axes which correspond to the variance (and hence power) in a particular direction in the data, each projection (or component) does not correspond to a discrete power band in the frequency domain. In fact, the components of the *eigen*spectrum formed from an SVD are not completely set before hand, (as is the case for an FFT-based analysis where each component corresponds to a particular frequency interval) but are **learned from the data**. In general, each component will overlap in the frequency domain to some extent. We will now explore SVD using a similar data set to that which we have just analyzed.

Load and plot a new set of ECG data using the following Matlab commands:

```
load X.dat
plot3ch(X)
```

This time you are not given the individual signal sources (mECG, fECG and noise), but rather three channels of (observed) ECG, all of which have a component of fetal ECG, maternal ECG and observation noise. The data in **X** is three channels of a 10-second ECG segment sampled at 256 Hz.

The function **plot3ch** produces two figures. The first plots the three ECG channels in the time domain. The second is a scatter plot of the three ECG channels. The scatter plot demonstrates how the data varies in one channel (as a function of the other channels). Note that most of the data is located near the origin (which is the mean for a zero-mean data set). Data points outside of the central cluster are mostly associated with the QRS peaks.

Now we will attempt to separate the maternal and fetal ECG using SVD. Recall Eq. 4 in chapter 15, ($\mathbf{Y} = \mathbf{USV}^T$), where $\mathbf{Y}$ (a real $M \times N$ matrix) can be decomposed into three other matrices, $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}^T$. $\mathbf{S}$ is a non-square diagonal matrix of singular

6

values whose elements are arranged in descending order of magnitude and represent the eigenvalues of $\mathbf{C} = \mathbf{Y} * \mathbf{Y^T}$, that is,

```
S =[    s1         0          0
        0          s2         0
        0          0          s3
        0          0          0
        ...        ...        ...
        0          0          0 ]
```

for N=3.

A plot of these diagonal elements is known as an *eigenspectrum*. The columns of **V** are the eigenvectors of $\mathbf{C} = \mathbf{Y} * \mathbf{Y}^T$ (with the lengths of the eigenvectors normalized to unity) and the matrix **U** is the matrix of projections of **Y** onto the eigenvectors of **C**. Perform SVD on the matrix of observations (*X*) using the Matlab function **svd** to create the matrices, **U**, **S** and **V**.

Plot the eigenvectors onto the observations *X* using the function **plot3dv**. Please note the following **important** points:

- The function **plot3dv** takes as input a column of the matrix *V* and the corresponding eigenvalue to scale the eigenvector. It may be called multiple times, once for each eigenvector.

- Before calling **plot3dv**, make sure your current (active) Matlab figure is the scatter plot produced earlier. This will ensure that the eigenvectors are superimposed on the scatter plot.

- No text response is necessary at this point, but please **be sure to save the matrix V and this Matlab figure for later comparisons**. In order to manipulate the three-dimensional figure later, you must save it in Matlab's .fig format. You can do this from the 'file' menu on the top left of the Matlab figure window - select 'Save As' and make sure that the file type is *.fig.

Recall that the matrix **U** is the matrix of projections of **Y** onto the eigenvectors of **C** [3]. The matrix **U** has a size of $M \times M$ but only the first N projections are meaningful because the dimensionality of the singular value decomposition space cannot be larger than the dimensionality of the data. All the columns after the first N columns are multiplied by the zeros in the **S** matrix!

Question 6 : *Plot the first three columns of the matrix **U**. For each of these projections, identify the dominant features (maternal, fetal, noise, or mixture). Indicate which projection is most representative of the fetal ECG. Also plot the eigenspectrum using **stem** and include this in your lab writeup.*

7

Note that this projection is no longer clinically meaningful. Traditional diagnosis from the ECG is performed by checking to see if lead specific thresholds of certain features are within certain bounds. By performing SVD we transform the data into another (non-clinical) 'space' in which lead specific thresholds have changed (depending on the structure of the data itself). Therefore if we wish to preserve the clinical information, we must invert the SVD transformation to project the ECG back into the original observation space. We can simultaneously extract the fetal ECG by modifying the **S** matrix before inverting the SVD transformation. Keep the eigenvalue corresponding to the output channel most representative of fetal ECG and set all other eigenvalues to 0. (The matrix is now singular and noninvertible, so we have a lossy transformation). Use the original **U** and **V** matrices along with the modified **S** to reconstruct the channels of fetal ECG.

Question 7 *Plot the three channels of your reconstructed fetal ECG signal. Describe the effects of the SVD reconstruction. Was it successful in extracting the fetal ECG?*

Note: Be sure to save the output of your SVD reconstruction, as you will need to compare it with the results from other techniques later in this lab.

2.4 Separating sources using ICA

In this section we will use a higher-order technique; *Independent Component Analysis* (ICA) [4] to perform the maternal-fetal-noise signal separation. Recall the mixing paradigm **observation = mixing matrix × underlying sources**, or

$$\mathbf{X^T} = \mathbf{AZ^T}. \tag{2}$$

Note that for dimensional consistency we have transposed the three channels of observations **X**, (originally M data points/rows by N channels/columns) and the three source column vectors **Z** (originally also $M \times N$).

The de-mixed estimate of the sources **Z** is therefore

$$\mathbf{\hat{Z}^T} = \mathbf{WX^T}. \tag{3}$$

where **W** is the de-mixing matrix and $\mathbf{W^{-1}}$ is the estimate of the mixing matrix **A**.

Perform ICA on the matrix of observations ($\mathbf{X^T}$) using the Matlab function *ica* to create the matrices *W* and $\mathbf{\hat{Z}}$. Save these matrices as they will be used in later computations.

Compute $\mathbf{W^{-1}}$, the estimate of the mixing matrix *A*. Please **be sure to save the matrix** $\mathbf{W^{-1}}$ **for later comparison.**

Make a scatter plot of the data (using **plot3ch**) and the three independent components along which the sources are projected (using **plot3dv**). (Note that **plot3dv** will take as

8

input a column of the matrix $\mathbf{W}^{-1}$ and can be called multiple times. No text response is necessary at this point, but please **be sure to save this Matlab figure for later comparisons**. In order to manipulate the three-dimensional figure later, you must save it in Matlab's .fig format. You can do this from the 'file' menu on the top left of the Matlab figure window - select 'Save As' and make sure that the file type is *.fig.

Question 8 *Plot the three columns of the matrix $\hat{\mathbf{Z}}$ (equivalently, the three rows of $\hat{\mathbf{Z}}^{\mathbf{T}}$). For each of the projections identify the dominant features (maternal, fetal, noise, or mixture). Indicate which projection is most representative of the fetal ECG.*

We can extract the fetal ECG by modifying the $\mathbf{W}^{-1}$ matrix. Keep the column of $\mathbf{W}^{-1}$ most representative of the fetal ECG and set the columns corresponding to the maternal and noise projections to 0. Use the $\hat{\mathbf{Z}}$ matrix along with the modified $\mathbf{W}^{-1}$ matrix to reconstruct the three channels of fetal ECG.

Question 9 *Plot and describe the effects of the ICA reconstruction. Was this successful in extracting the fetal ECG?*

Note: Be sure to save the output of your ICA reconstruction, as you will need to compare it with the results from other techniques later in this lab.

2.5   Comparisons

Question 10 *Load the scatter plots created in Sections 2.3 and 2.4 showing the observations $\mathbf{X}$ plotted along two different sets of axes. Rotate and compare the two figures. Now quantitatively compare the two sets of axes. Use the Matlab function **dot** to do pairwise comparisons of the angles of the vectors represented by the columns of $\mathbf{V}$. Do the same with the columns of $\mathbf{W}^{-1}$. Note that when you are done you will have three scalar values for each axis. What is the major (single) difference between PCA and ICA as evidenced by these angles? Also calculate the lengths of each of axis using the Matlab function **norm**. How do the lengths compare?*
***Hint**: Don't try to interpret subtle differences between the numerical values of the dot products and vector lengths; focus on general differences.*

One essential question we must ask is *how well did the signal/noise separation work?* The only exact way to assess performance is if we know what exactly is the signal and what exactly is the noise. Since we are using simulated data, we have perfect knowledge of the desired signal (fECG) and the noise (mECG+noise). For the data matrix $\mathbf{X}$ that you have been working with, the desired fECG is given in **fecg2.dat**, which is a scaled version of **fecg1**.

9

Question 11 *Compare the results of Wiener filtering, SVD, and ICA (from questions 5, 7, & 9) with a plot of the ideal fetal ECG (**fecg2**). Please describe qualitatively the relative effectiveness of each filtering technique. Which one did the best? Which one did the worst? Note that there are three channels of filtered fetal ECG for the SVD and ICA outputs in comparison to the single channel in the Wiener filter.*

Question 12 *Now quantitatively evaluate the effectiveness of each filtering technique by using the correlation coefficient to compare the ideal output (**fecg2**) with the actual output from each technique. For PCA and ICA, you can choose one channel from each output. Use the Matlab function **corrcoef**. Please include these results in your lab writeup. Which method produced the cleanest fetal ECG signal, and which method performed the worst?*

Question 13 *Compare and contrast the three filtering techniques explored in this lab. Describe the advantages and limitations of each technique. (Suggested length: one paragraph).*

Question 14 *What is the most important thing that you learned from this lab exercise?*

Question 15 *What did you like/dislike the most about this lab exercise?*

References

[1] Oostendorp T, vanOosterom A, H.W. J. The fetal ECG throughout the second half of gestation. Clin Phys Physiol Meas 1989;10(2):147–160.

[2] Richter M, Schreiber T, Kaplan DT. Fetal EEG extraction with nonlinear phase space projections. IEEE Trans Bio Med Eng Jan 1998;45:133–137.

[3] Golub GH, Van Loan CF. Matrix Computations. 2nd edition. Oxford: North Oxford Academic, 1983.

[4] Cardoso JF. High-order contrasts for independent component analysis. Neural Comput 1999;11(1):157–192. ISSN 0899-7667.