1 Random Projection Question

Assume that the d dimensional vector $\mathbf{u}$ is obtained by setting each of its coordinates to either $+1$ or -1 based on coin flips. That is, for each $i \in [d]$, $\mathbf{u}[i] = \begin{cases} +1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$.
Now set $y_t = \mathbf{x}_t^\top \mathbf{u}$ for every $t \in [n]$.

Question: For some $t, s \in [n]$, what is:

1. $\mathbb{E}[y_t - y_s]$?
2. $\mathbb{E}[(y_t - y_s)^2]$?

2 Canonical Correlations Analysis Question

In CCA we think of data $\mathbf{x}_t$ as coming in pairs $\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}'_t \end{bmatrix}$. where say $\mathbf{x}_t$ is say d dimensional and $\mathbf{x}'_t$ is d' dimensional. Accordingly we can think of data matrix as $[X, X']$. That is n pairs of points. For example, we might get pictures from two different camera angles, or we might obtain audio and video components for a video recording. Our goal in CCA is to extract the common information amongst the two views from any one of the views.

To this end, we start with linear dimensionality reduction with such two views. For this handout, we consider one dimensional projections in each of the views. That is, we are interested in reducing the d and d' dimensional data in the two views into single numbers in each view by a linear transformation. To this end, let the one dimensional projection in view I be given by numbers $y_1, \ldots, y_n$ where

$$y_t = \mathbf{w}_1^\top \mathbf{x}_t \;\; \& \;\; y'_t = \mathbf{v}_1^\top \mathbf{x}'_t$$

2.1 Question 1

You want to find directions $\mathbf{w}_1$ and $\mathbf{v}_1$ that maximize the correlation coefficient given by:

$$\frac{1}{n}\sum_{t=1}^{n} \frac{\left(y_t - \frac{1}{n}\sum_{t=1}^{n} y_t\right) \cdot \left(y'_t - \frac{1}{n}\sum_{t=1}^{n} y'_t\right)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \frac{1}{n}\sum_{t=1}^{n} y_t\right)^2}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y'_t - \frac{1}{n}\sum_{t=1}^{n} y'_t\right)^2}}$$

Why is this problem equivalent to finding directions $\mathbf{w}_1$ and $\mathbf{v}_1$ that maximize

$$\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \frac{1}{n}\sum_{t=1}^{n} y_t\right) \cdot \left(y'_t - \frac{1}{n}\sum_{t=1}^{n} y'_t\right)$$

subject to condition:

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_t - \frac{1}{n}\sum_{t=1}^{n} y_t\right)^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y'_t - \frac{1}{n}\sum_{t=1}^{n} y'_t\right)^2} = 1$$