Example 10.1. Correlation Analysis. Let \( x_1, x_2, \ldots, x_n \) be a time series (e.g. an action potential over time) that we store in a vector \( x = (x_1, x_2, x_3, \ldots, x_n)^\top \in \mathbb{R}^n \). For simplicity, we assume that \( x \) has been pre-processed version such that \( m(x) = 0 \), and \( m_2(x) = 1 \). Let \( y \) be another measurement, also with \( m(y) = 0 \) and \( m_2(y) = 1 \).

Reminder: \( \text{corr}(x, y) \) tells us how similar \( x \) and \( y \) are (note: \( \text{corr}(x, y) = \langle x, y \rangle \) because of the assumed normalization).

Time series that were measured at different locations often exhibit a time shift (lag), so \( x \) and \( y \) would only be similar, if we shifted one of them in time. This is measured by the cross-correlation.

\[
R_{xy}(\tau) := \sum_{i=1}^{n-\tau} x_i y_{i+\tau} \quad \text{for } \tau \in \{0, 1, \ldots, n-1\}. \tag{10.1}
\]

For fixed \( x \) and \( y \) we get one value for each \( \tau \in \{0, \ldots, n\} \), which measures how similar \( x \) is to a \( y \) that was shifted by \( \tau \) steps to the left. A shift to the right we can define in a similar way, or we just exchange the role of elements and shift \( x \) to the left:

\[
R_{xy}(\tau) := R_{yx}(-\tau) \quad \text{for } \tau \in \{-1, -2, \ldots, -n + 1\}. \tag{10.2}
\]

\( R_{xy}(\tau) \) is a function that we can plot: its maximum tells us at which offset \( x \) and \( y \) are most similar.
We can write $R_{xy}$ as a matrix-vector operation: let
\[ s_\tau(y) = (y_\tau, y_{\tau+1}, \ldots, y_n, 0, \ldots, 0)^\top \]
be the (linear) function that shifts $y$ to the left. Its matrix is a shifted version of the identify matrix, e.g. $S_\tau = \begin{pmatrix} 0 & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$ for $\tau = 2$.\[ R_{xy}(\tau) := \sum_{i=1}^{n-\tau} x_i y_{i+\tau} = \sum_{i=1}^{n} x_i [s_\tau(y)]_i = x^\top S_\tau y \quad (10.3) \]

For periodic signals, we can use cyclic shift:
\[ s_\tau(x) = (y_\tau, y_{\tau+1}, \ldots, y_n, y_1, \ldots, y_{\tau-1})^\top. \]
This has the advantage that $\hat{s}$ keeps its normalization, we always compare signals of the same length.

**Special Case: Autocorrelation** In the case where $x = y$ we call $R_{xx}(\tau)$ the autocorrelation function of $x$. Its maximum will be at $\tau = 0$, and we have $R_{xx}(0) = ||x||^2 = 1$ there. However, other offsets are also interesting: if $R_{xx}(\tau)$ is large for $\tau \neq 0$, then $x$ is similar to a shifted version of itself. For example, if $x$ is a periodic signal (potentially with noise), then $R_{xx}(\tau)$ peaks at each period.

**Expressing vectors as weighted combinations of other vectors**

**Reminder from high school geometry.** Projecting a point to a line:
Let $v \in \mathbb{R}^n$, $v \neq 0$, be a vector, and let $L_v = \{x : x = \lambda v \text{ for } \lambda \in \mathbb{R}\}$ be the line going through the origin with in the direction of $v$.

Denote by $y = p(x)$ the point on $L_v$ closest to $x$. 
LECTURE 10. CORRELATION ANALYSIS / PROJECTIONS

- 1) \( y \in L_v \)
- 2) \( \|x - y\|^2 \) is minimal (equivalent, to "\( \|x - y\| \) is minimal").

From these two facts, we can find the distance of \( x \) to \( L_v \):

\[
\min_{y \in L_v} \|x - y\|^2 = \min_{\lambda \in \mathbb{R}} \|x - \lambda v\|^2 =: D(\lambda) \tag{10.4}
\]

Note \( D(\lambda) \) is a (non-linear) function of a single real-valued parameter, \( \lambda \):

\[
D(\lambda) = \langle x - \lambda v, x - \lambda v \rangle \tag{10.5}
\]

\[
= \langle x, x \rangle - \langle v, \lambda x \rangle - \langle \lambda x, v \rangle + \langle \lambda v, \lambda v \rangle \tag{10.6}
\]

\[
= \langle x, x \rangle - 2\lambda \langle x, v \rangle + \lambda^2 \langle v, v \rangle \tag{10.7}
\]

A necessary condition for minimum is \( \frac{d}{d\lambda} D(\lambda) = 0 \); so

\[
0 = \frac{d}{d\lambda} D(\lambda) = -2\langle x, v \rangle + 2\lambda \|v\|^2 \tag{10.8}
\]

so

\[
\lambda = \frac{\langle x, v \rangle}{\|v\|^2} \tag{10.9}
\]

For any \( x \), the point on \( L \) closest to \( x \) is \( p(x) = \frac{\langle x, v \rangle}{\|v\|^2} v \).

**Lemma 10.2.** For any line \( L_v \), let \( p : \mathbb{R}^n \to \mathbb{R}^n \) be the function that projects a points \( x \) to the point on the line closest to \( x \). Then \( p \) is a linear function with matrix \( P := \frac{1}{\|v\|^2} vv^\top \in \mathbb{R}^{n \times n} \).

**Proof.** We saw above that \( p(x) = \frac{\langle x, v \rangle}{\|v\|^2} v \). We only have to check that \( p(x) = Px \) with \( P \) defined as above.

\[
P x = \left( \frac{1}{\|v\|^2} vv^\top \right) x = \frac{1}{\|v\|^2} v \left( v^\top x \right) = \frac{\langle x, v \rangle}{\|v\|^2} v = p(x) \tag{10.10}
\]

\[\square\]

**Remark.** Since for specifying the line the length of \( v \) doesn’t matter, so often one uses a normalized \( v \) (i.e. \( \|v\| = 1 \)), then the projection simplifies to \( p(x) = \langle x, v \rangle v \), and \( P = v^\top v \).

**Remark.** We can interpret \( p(x) \) as the best approximation of \( x \) which only using a multiple of \( v \). \( \frac{\langle x, v \rangle}{\|v\|^2} \) then computes "How much of \( v \) is in \( x \)?"
Measuring Properties by Projection

**Example 10.3.** Let \( s = (s_1, \ldots, s_n)^\top \) be the values of a time series measured at points of time \( t = (t_1, \ldots, t_n) \)

- Let \( v = (1, 1, \ldots, 1)^\top \) be a constant signal.

\[
P_t(s) = \langle s, v \rangle \frac{1}{\|v\|^2} = \langle s, v \rangle \frac{1}{\sum_{i=1}^{n} t_i} = \begin{pmatrix} m(s) \\ m(s) \\ \vdots \\ m(s) \end{pmatrix}
\]

The best approximation of \( s \) by a constant signal is the mean function.

- Let \( v = (t_1, t_2, \ldots, t_n)^\top \) be the linear (identity) signal.

\[
P_v(s) = \langle s, v \rangle \frac{1}{\|v\|^2} = \frac{\sum_{i=1}^{n} s_i t_i}{\sum_{i=1}^{n} t_i^2} = \begin{pmatrix} c t \end{pmatrix}
\]

Coefficient \( c \) is the same as in least-squares regression, when fitting a linear function \( f(t) = ct \) to the measurements \( s_i \).

- Let \( v_\omega = \left( \sin(\omega t_1), \sin(\omega t_2), \ldots, \sin(\omega t_n) \right)^\top \) be a sine wave of period \( \omega \). \( P_{v_\omega}(s) = \frac{\langle s, v_\omega \rangle}{\|v_\omega\|} v \) is the best sinusoidal fit with period \( \omega \).

- For any direction \( v \) and signal \( s \): \( \|P_{v_\omega}(s) - s\| \) measures how good the fit is.

This is in absolute terms: if \( s \) is measured in \( g \) the difference is typically larger than if it is measured in tons. Often more interesting: relative goodness of fit:

\[
\frac{\|P_{v_\omega}(s)\|}{\|s\|} = \frac{1}{\|s\|} \frac{\|\langle s, v \rangle v\|}{\|v\|^2} = \frac{\|\langle s, v \rangle\|}{\|s\|\|v\|^2} = \frac{|\langle s, v \rangle|}{\|s\|\|v\|} = |\text{corr}(s, v)|
\]

Note: other goodness-of-fit measures can be motivated in the similar way, e.g. \( \chi^2 \) statistics, ANOVA...
Lemma 10.4. Some properties of projections:
1) Projecting twice to $v$ is the same as doing it just once: $p_v(p_v(x)) = p_v(x)$.
2) If $v$ and $w$ be orthogonal non-zero vectors, then $p_v(p_w(x)) = 0$.

Let $y = x - p_v(x)$ ("what remains from $x$ after removing all of $v$ that $x$ contains"). Then
3) $p_v(y) = 0$, and
4) $y$ is orthogonal to every vector in $L_v$, i.e. $y \perp z$ for every $z \in L_v$.

Proof. Side-Computations:
\[ \langle p_v(x), v \rangle = \left\langle \frac{\langle x, v \rangle}{\|v\|^2} v, v \right\rangle = \frac{\langle x, v \rangle}{\|v\|^2} \|v\|^2 = \langle x, v \rangle. \] (10.13)

For $w \perp v$:
\[ \langle p_v(x), w \rangle = \left\langle \frac{\langle x, v \rangle}{\|v\|^2} v, w \right\rangle = \frac{\langle x, v \rangle}{\|v\|^2} \langle v, w \rangle = 0. \] (10.14)

Proof of 1)
\[ p_v(p_v(x)) = \frac{\langle p_v(x), v \rangle}{\|v\|^2} v = \frac{\langle x, v \rangle}{\|v\|^2} v = p_v(x) \]

Proof of 2) for $w \perp v$:
\[ p_w(p_v(x)) = \frac{\langle p_v(x), w \rangle}{\|w\|^2} w = 0 \]

Proof of 3) for $y = x - p_v(x)$:
\[ \langle y, v \rangle = \langle x - p_v(x), v \rangle = \langle x, v \rangle - \langle p_v(x), v \rangle = \langle x, v \rangle - \langle x, v \rangle = 0. \]

Proof of 4) for any $\lambda \in \mathbb{R}$:
\[ \langle y, \lambda v \rangle = \lambda \langle y, v \rangle = 0. \]

Example 10.5. Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$.
\[ v_1 = e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad p_1(x) = \langle x, e_1 \rangle e_1 = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}, \quad p_1(p_1(x)) = p_1(\begin{pmatrix} x_1 \\ 0 \end{pmatrix}) = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}, \]
\[ x - p_1(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ x_2 \end{pmatrix}, \quad \left\langle \begin{pmatrix} 0 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \right\rangle = 0 \]
• \( v = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \). \( p(x) = \frac{1}{\|v\|^2} \langle v, e_2 \rangle v = \frac{1}{2} (x_1 + x_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \left( \frac{x_1 + x_2}{x_1 + x_2} \right) \).

\[ x - p(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \left( \frac{x_1 + x_2}{x_1 + x_2} \right) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \left( \frac{x_1 - x_2}{2} \right) \left( -\frac{x_1 + x_2}{2} \right) \).

\[ \langle \left( \frac{x_1 + x_2}{x_1 + x_2} \right), \left( \frac{x_1 - x_2}{x_1 + x_2} \right) \rangle = \frac{1}{4} \left( \begin{pmatrix} x_1 + x_2 \\ x_1 + x_2 \end{pmatrix}, \begin{pmatrix} x_1 - x_2 \\ -x_1 + x_2 \end{pmatrix} \right) \]

\[ = \frac{1}{4} \left[ (x_1 + x_2)(x_1 - x_2) + (x_1 + x_2)(x_2 - x_1) \right] = 0. \]