1 Gradient of the norm function

Recall that the gradient of a differentiable function \( f : \mathbb{R}^n \to \mathbb{R} \) at a point \( x \in \mathbb{R}^n \) is defined to be the vector

\[
\nabla f(x) = \begin{bmatrix}
\frac{\partial f}{\partial x_1}(x) \\
\vdots \\
\frac{\partial f}{\partial x_n}(x)
\end{bmatrix}.
\]

The first-order Taylor expansion of \( f \) near \( x \) is given by

\[
\hat{f}_1(z) = f(x) + \nabla f(x)^T(z - x).
\]

This function is affine: that is, a linear function plus a constant offset. If \( z \) is near \( x \), then \( \hat{f}_1(z) \) is very near \( f(z) \). Find the gradient of the function \( f(x) = \|x\| \).

2 Some standard time-series models

In some contexts, a discrete-time signal is called a time series. The study of time series predates the extensive study of linear state-space systems, and is used in many fields. Let \( u \) and \( y \) be two time series, which we will think of as the input and output, respectively.

(a) The relation (or time-series model)

\[
y(k) = a_0 u(k) + a_1 u(k - 1) + \cdots + a_r u(k - r)
\]

is called a moving-average (MA) model. Since the output at time \( k \) is a weighted average of the previous \( r \) inputs, we can think of the output as the average of the inputs in a moving window. Express this model as a linear dynamical system with input \( u \), output \( y \), and state

\[
x(k) = \begin{bmatrix}
u(k - 1) \\
\vdots \\
u(k - r)
\end{bmatrix}.
\]

(b) Another time-series model is

\[
y(k) = u(k) + b_1 y(k - 1) + \cdots + b_p y(k - p).
\]

This model is called an autoregressive (AR) model, since the current output is a linear combination of the current input, and some previous values of the output. Express this model as a linear dynamical system with input \( u \), output \( y \), and state

\[
x(k) = \begin{bmatrix}
y(k - 1) \\
\vdots \\
y(k - p)
\end{bmatrix}.
\]
(c) A third widely used model is the autoregressive, moving-average (ARMA) model, which combines the MA and AR models:

\[ y(k) = a_0 u(k) + \cdots + a_r u(k - r) + b_1 y(k - 1) + \cdots + b_p y(k - p). \]

Express this model as a linear dynamical system with input \( u \) and output \( y \) (you can choose the state; there are many possible choices, and not all choices have the same dimension).

3 Some linear functions associated with a convolution system

Suppose that \( u \) and \( y \) are discrete-time scalar signals related via convolution:

\[ y(t) = \sum_{\tau = -\infty}^{+\infty} h(t - \tau) u(\tau), \quad t \in \mathbb{Z}, \]

where \( (h(t) : t \in \mathbb{Z}) \) is a known discrete-time scalar signal. You may assume that the system is causal: that is, \( h(t) = 0 \) for all \( t < 0 \).

(a) Suppose \( u(t) = 0 \) for all \( t < 0 \). Define the vectors

\[ \vec{u} = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(T) \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(T) \end{bmatrix}. \]

Thus, \( \vec{u} \) and \( \vec{y} \) are the first \( T + 1 \) values of the input and output signals, respectively. Find the matrix \( G \in \mathbb{R}^{(T+1) \times (T+1)} \) such that \( \vec{y} = G \vec{u} \). The matrix \( G \) describes the linear mapping from (a segment of) the input sequence to (a segment of) the output sequence; \( G \) is called the input/output or Toeplitz matrix of size \( T + 1 \) associated with the convolution system.

(b) Now suppose that \( u(t) = 0 \) for all \( t < 0 \) and \( t > T \). Define the vectors

\[ \vec{u} = \begin{bmatrix} u(T) \\ u(T - 1) \\ \vdots \\ u(0) \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} y(T) \\ y(T + 1) \\ \vdots \\ y(2T) \end{bmatrix}. \]

Thus, \( \vec{u} \) is the input to the system, and \( \vec{y} \) is (a segment of) the future output of the system. Find the matrix \( H \in \mathbb{R}^{(T+1) \times (T+1)} \) such that \( \vec{y} = H \vec{u} \). The matrix \( H \) describes the linear mapping from the input sequence to (a segment of) the future output sequence; \( H \) is called the Hankel matrix of size \( T + 1 \) associated with the convolution system.
4 Counting paths in an undirected graph
Consider an undirected graph with \( n \) nodes, and no self loops. Let \( A \in \mathbb{R}^{n \times n} \) be the node-adjacency matrix, which is defined such that
\[
A_{ij} = \begin{cases} 
1 & \text{there is an edge between nodes } i \text{ and } j, \\
0 & \text{otherwise.} 
\end{cases}
\]
Note that \( A = A^T \) because the graph is undirected, and \( A_{ii} = 0 \) since there are no self loops. Give an interpretation of \((A^p)_{ij}\) (that is the \((i,j)\)-entry of \(A^p\)) for \( p \in \mathbb{N} \).

5 Memory of a linear, time-invariant system
Suppose an input signal \( (u_t : t \in \mathbb{Z}) \), and an output signal \( (y_t : t \in \mathbb{Z}) \) are related by a convolution operator:
\[
y_t = \sum_{\tau=1}^{M} h_{\tau} u_{t-\tau},
\]
where \( h = (h_1, \ldots, h_M) \) are the impulse-response coefficients of the convolution system. (Convolution systems are also called linear, time-invariant systems.) If \( h_M \neq 0 \), then \( M \) is called the memory of the system. You are given the input and output signals for \( t = 1, \ldots, T \):
\[
u_1, \ldots, u_T \quad \text{and} \quad y_1, \ldots, y_T.
\]
However, you do not know \( u_t \) or \( y_t \) for \( y < 1 \) or \( t > T \), and you do not know the impulse response, \( h \).

(a) Explain how to find the smallest value of \( M \), and a corresponding impulse response \((h_t : t = 1, \ldots, M)\) that is consistent with the given data. You may assume that \( T > 2M \).

(b) Apply your method to the data in \texttt{lti\_memory\_data.m}. Report the value of \( M \) that you find.

*Hint.* The function \texttt{toeplitz} may be useful.