# Homework 13

## This homework is due on Saturday, April 27, 2024 at 11:59PM.

**1. PCA and LS Prediction**

In this problem, you will be helping to analyze some clinical data for patients using PCA and least squares. Suppose the following is a data matrix consisting of scaled *zero-mean* patient data. The patients (i.e. data points) are the *columns*, while the rows from top to bottom represent scores for the height, weight, and blood pressure of the 4 patients.

$$A = \begin{bmatrix} -3 & 0 & 2 & 1 \\ -1 & -2 & 0 & 3 \\ -1 & -1 & 1 & 1 \end{bmatrix} \tag{1}$$

Using this information, we would like to predict a risk index for heart disease for each patient (negative if low-risk and positive if high risk). But unfortunately, there appears to be some noise mixed into the data. Let's try to mitigate this by *performing PCA to embed the training data into a lower dimension first*.

(a) Let us perform PCA on $A$ to embed the data into a single dimension. Suppose you know that $\sigma_1 = \sqrt{8}$ and $\vec{v}_1 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}^T$ from the first element of $\Sigma$ and first column of $V$ for the SVD of $A$. **What is the first principal component $\vec{u}_1$ of $A$? Show all of your work.**

(b) Regardless of your answer to the previous question, suppose $\vec{u}_1 = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}^T$. **Project your three-dimensional data points from $A$ into one-dimensional values. Express your answer as the vector $\vec{z} \in \mathbf{R}^{1\times 4}$, and show all your work.**

(c) Now, suppose that for this training data we have known risk index scores already (with a scalar index for each patient):

$$\vec{b} = \begin{bmatrix} -3 & -1 & 4 & 6 \end{bmatrix} \tag{2}$$

We would like to be able to predict risk scores. Let's try to use least squares on our scenario. In one dimension, we can set up our least squares problem using data $\vec{d} = \vec{z}^T$ and target $\vec{s} = \vec{b}^T$ to estimate a scalar parameter $p_{\text{LS}}$. **Perform least squares on the system $\vec{d}p_{\text{LS}} = \vec{s}$ to estimate $p_{\text{LS}}$, and show your work.**

**2. Movie Ratings and PCA**

Recall from the lecture on PCA that we can think of movie ratings as a structured set of data. For every person $i$ and movie $j$, we have that person's rating $R_{i,j}$ (thought of as a real number).

Suppose that there are $m$ movies and $n$ people. Let's think about arranging this data into a big $n \times m$ matrix $R$ with people corresponding to rows and movies corresponding to columns. So the entry in the $i$-th row and $j$-th column should be $R_{i,j}$ above. Note that this is organized differently from how it was in lecture. Each row corresponds to a unique person and each column to a unique movie.

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix} \tag{3}$$

(a) Suppose we believe that there is actually an underlying pattern to this rating data and that a separate study has revealed that every movie is characterized by a set of characteristics: say action and comedy. This means that every movie $j$ has a pair of numbers $a_j$ (for action) and $c_j$ (for comedy) that define it. At the same time, every person $i$ has a pair of sensitivities $f_i$ and $g_i$ that define their preferences for action vs. comedy in movies respectively. A person $i$ will rate the movie $j$ as $R_{i,j} = f_i a_j + g_i c_j$.

If we arrange the sensitivities into a pair of $n$-dimensional vectors $\vec{f}, \vec{g}$ for our group of $n$ people, and the movie characteristics into a pair of $m$-dimensional vectors $\vec{a}, \vec{c}$ for our group of $m$ movies, **use outer products to express the rating matrix $R$ in terms of these vectors $\vec{f}, \vec{g}, \vec{a}, \vec{c}$.**

(b) Now suppose that we want to discover the underlying nature of movies from the data $R$ itself.

Suppose for this part that you have four observed rating data vectors (corresponding to four different movies being rated by six individuals).

All of the movie data vectors just happened to be multiples of the following 6-dimensional vector $\vec{w} = \begin{bmatrix} 2 \\ -2 \\ 3 \\ -4 \\ 4 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{w}\| = 7$.)

You arrange the movie data vectors as the columns of a matrix $R$ given by:

$$R = \begin{bmatrix} | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} \\ | & | & | & | \end{bmatrix} \tag{4}$$

You want to perform PCA (for movies) using the SVD of the matrix $R$ to better understand the pattern in your data.

The first "principal component vector" is a <u>unit vector</u> that tells which direction we would want to project the columns of $R$ onto to get the best rank-1 approximation for $R$.

**Find this first principal component vector of the columns of $R$ to explain the nature of your movie data points.**

*(HINT: You don't need to actually compute any SVDs in this simple case. Also, be sure to think about what size vector you want as the answer. Don't forget that you want a unit vector!)*

(c) Suppose that we now collected two more data points (corresponding to two more movies being rated by the same set of six people, i.e. two additional columns for our matrix) that are multiples of a different vector $\vec{p}$ where:

$\vec{p} = \begin{bmatrix} 6 \\ 3 \\ -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{p}\| = 7$ and that $\vec{p}^\top \vec{w} = 0$.)

We augment our ratings data matrix with these two new data points to get:

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix} \tag{5}$$

Find the first two principal components corresponding to the nonzero singular values of $R$. This is what we would use to best project the movie data points onto a two-dimensional subspace.

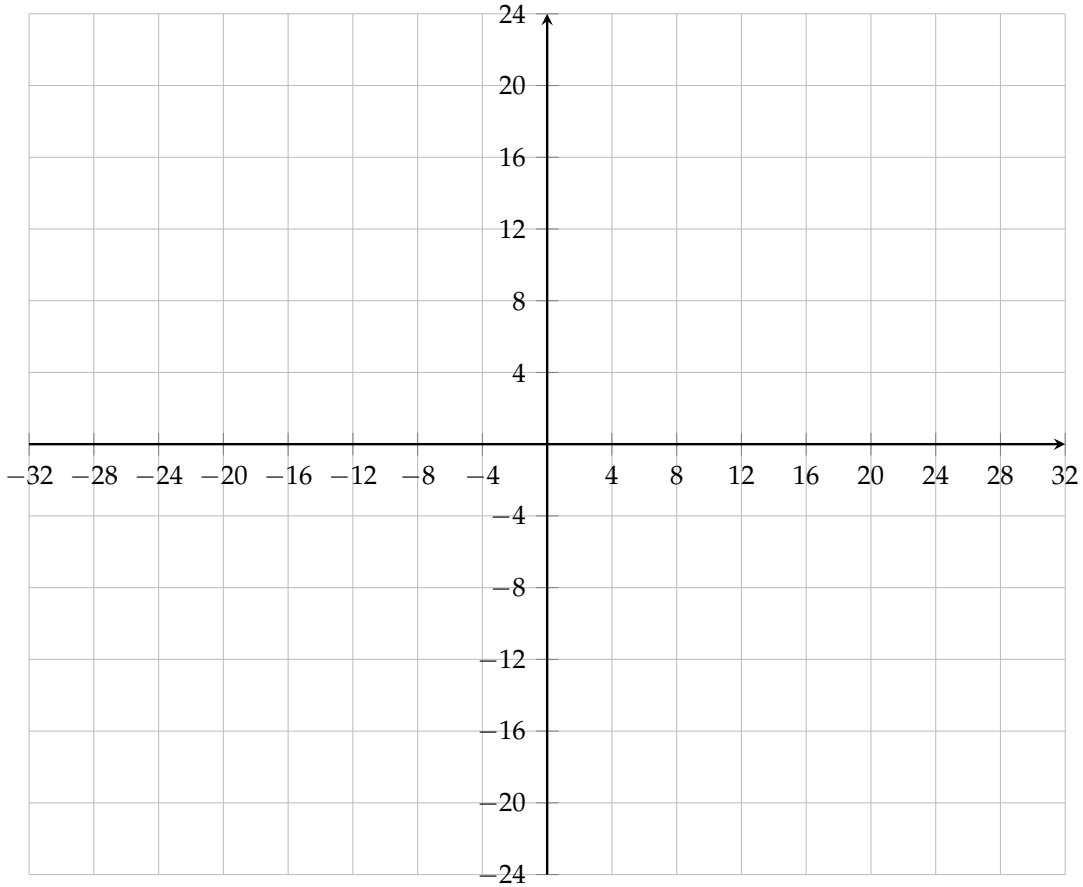
**What is the first principal component vector? What is the second principal component vector? Justify your answer.** *(HINT: Think about the inner product of $\vec{w}$ and $\vec{p}$ and what that implies for being able to appropriately decompose $R$. Again, very little computation is required here.)*

(d) In the previous part, you had

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix}$$

with $\|\vec{w}\| = 7$ and $\|\vec{p}\| = 7$, satisfying $\vec{p}^\top \vec{w} = 0$.

**Plot the movie data points $\vec{r}_i$ (for all $i$, where $\vec{r}_i$ denotes the $i$-th column of $R$) projected onto the first and second principal component vectors of the columns of R.** The horizontal axis should reflect the coordinate along the first principal component, and the vertical axis should reflect the coordinate along the second principal component. **Label each point, and the axes. Remember that principal component vectors are normalized**.



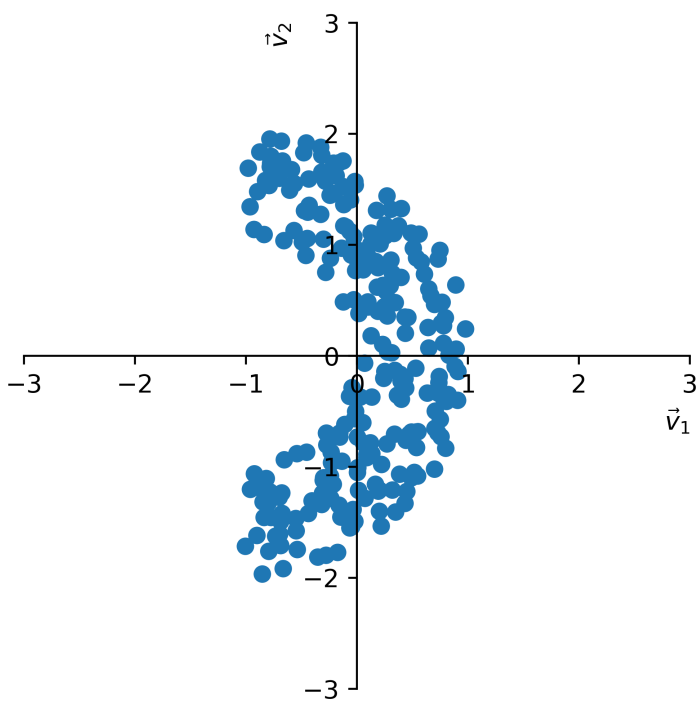

 

**3. PCA Plots**

In each plot below, some $d$-dimensional data is projected onto two unit vectors. The $x$-coordinate is the projection onto the first vector (written as $\vec{v}_1$), and the $y$-coordinate is the projection onto the second vector (written as $\vec{v}_2$). Mathematically, we can say that we have some data matrix $D = \begin{bmatrix} \vec{d}_1 & \vec{d}_2 & \cdots & \vec{d}_n \end{bmatrix}$ where each $\vec{d}_i \in \mathbb{R}^d$. We then project each $\vec{d}_i$ onto the column space of the matrix $V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$ and plot the projection coefficients below. We say that a plot is "valid" if $\vec{v}_1$ could be the first principal component.

*(HINT: Note that the mean or "center of mass" of the data points is the origin, (0, 0), in all of the plots. The procedure for this problem is very similar to the procedure in Discussion 14B, where you were similarly asked to judge the "validity" of scatterplots.)*

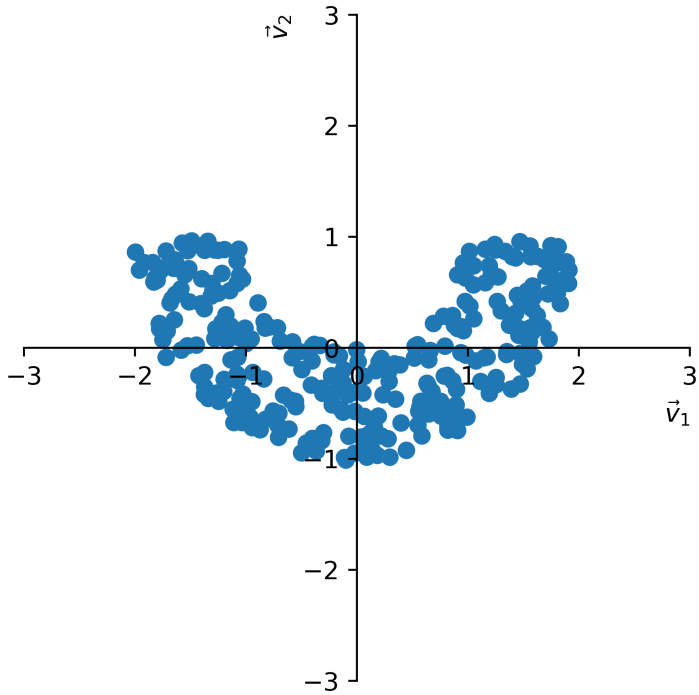
(a) **Is the following plot valid?**




| | |
|---|---|
| ◯ | Valid |
| ◯ | Invalid |

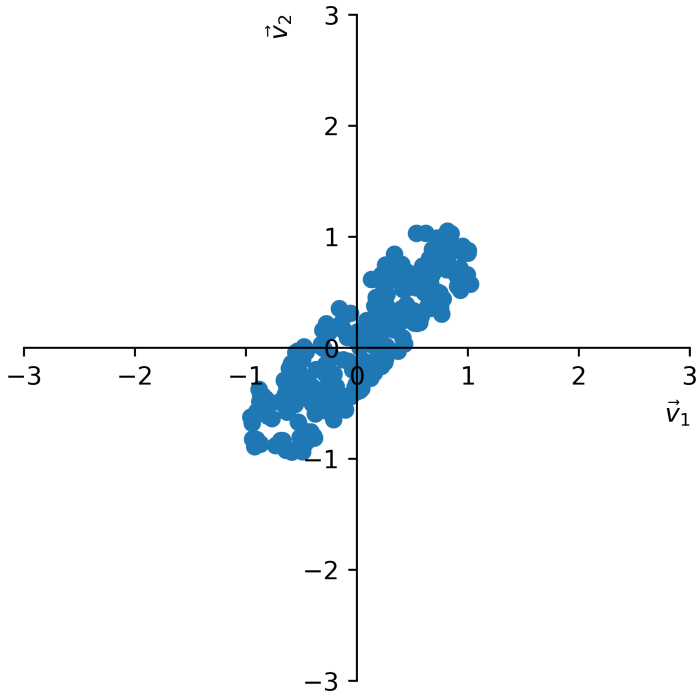
  

(b) **Is the following plot valid?**



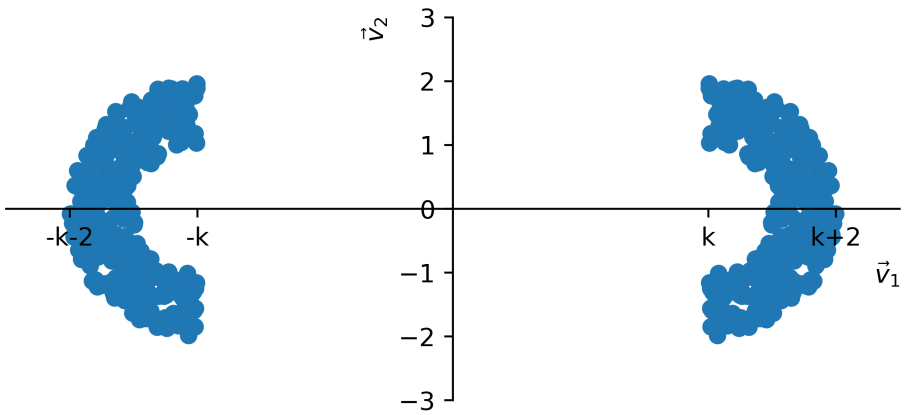

| | |
|---|---|
| ◯ | Valid |
| ◯ | Invalid |

(c) **Is the following plot valid?**




| | |
|---|---|
| ◯ | Valid |
| ◯ | Invalid |

(d) **Is the following plot valid?** Assume that $k \to \infty$.




$k \to \infty$

| | |
|---|---|
| ◯ | Valid |
| ◯ | Invalid |

**4. (OPTIONAL) Linearization of a Scalar System**

In this question, we linearize the scalar differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \sin(x(t)) + u(t) \tag{6}$$

around equilibria, discretize it, and apply feedback control to stabilize the resulting system.

(a) The first step is to find the equilibria that we will linearize around. Recall that equilibria are the values of $(x, u)$ such that $\frac{\mathrm{d}}{\mathrm{d}t}x(t) = 0$. Suppose we want to linearize around equilibria $(x^\star, u^\star)$ where $u^\star = 0$. **Sketch** $\sin(x)$ **for** $-4\pi \leq x \leq 4\pi$ **and intersect it with a horizontal line at** $0$**. Then, argue why** $x_m^\star = m\pi$ **and** $u^\star = 0$ **are equilibria of the system** (6)**.**

(b) Linearize system (6) around the equilibrium $(x_0^\star, u^\star) = (0, 0)$. **What is the resulting linearized scalar differential equation for** $\delta x(t) = x(t) - x_0^\star = x(t) - 0$**, involving** $\delta u(t) = u(t) - u^\star = u(t) - 0$**?**

(c) Given an arbitrary, continuous linear system as in

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \lambda x(t) + bu(t) \tag{7}$$

discretizing it into intervals of $\Delta$ gives the discrete-time system

$$x[i+1] = e^{\lambda\Delta}x[i] + \frac{b(e^{\lambda\Delta} - 1)}{\lambda}u[i] \tag{8}$$

**Using this result, discretize the approximate linear system. Is the (approximate) discrete-time system stable?**

(d) Suppose for the <u>linearized discrete-time system</u> that you found in the previous part, we apply the feedback law

$$\delta u[i] = -k(\delta x[i] - x^\star).$$

**For what range of** $k$ **values would the resulting linearized discrete-time system be stable?** Your answer will depend on $\Delta$.

**5. (OPTIONAL) Linearization to Understand Amplification**

Linearization isn't just something that is important for control, robotics, machine learning, and optimization — it is also one of the standard tools used across different hardware disciplines, including circuits.

The circuit below is a voltage amplifier, where the element inside the box is a bipolar junction transistor (BJT). You do not need to know what a BJT is to do this question.

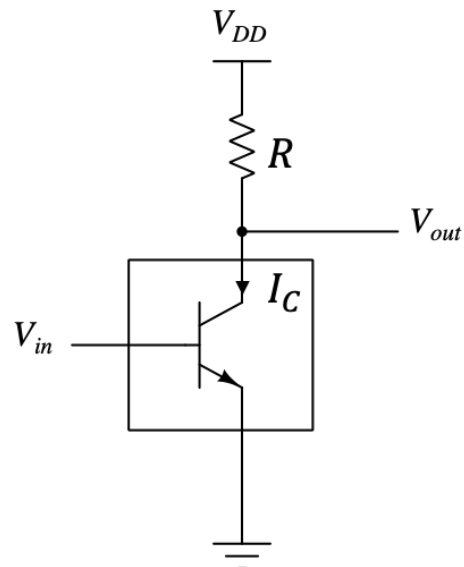



**Figure 1:** Voltage amplifier circuit using a BJT

The BJT in the circuit can be modeled quite accurately as a nonlinear, voltage-controlled current source, where the collector current $I_C$ is given by:

$$I_C(V_{\text{in}}) = I_S \cdot e^{\frac{V_{\text{in}}}{V_{\text{TH}}}}, \tag{9}$$

where $V_{\text{TH}}$ is the thermal voltage. We can assume that $V_{\text{TH}} = 26\,\text{mV}$ at room temperature. $I_S$ is a constant we are not providing because we want you to find ways of eliminating it in favor of other quantities whenever possible.

The goal of this circuit is to pick a particular point $(V_{\text{in}}^\star, V_{\text{out}}^\star)$ so that any small variation $\delta V_{\text{in}}$ in the input voltage $V_{\text{in}}$ can be amplified to a relatively larger variation $\delta V_{\text{out}}$ in the output voltage $V_{\text{out}}$. In other words, if $V_{\text{in}} = V_{\text{in}}^\star + \delta V_{\text{in}}$ and $V_{\text{out}} = V_{\text{out}}^\star + \delta V_{\text{out}}$, then we want the magnitude of the 'amplification gain' given by $\left|\frac{\delta V_{\text{out}}}{\delta V_{\text{in}}}\right|$ to be large. We're going to investigate this amplification using linearization.

(NOTE: In this problem, $\delta V$ is single variable indicating a small variation in $V$, not $\delta \times V$.)

(a) **Write a symbolic expression for $V_{\text{out}}$ as a function of $I_C$, $V_{\text{DD}}$ and $R$ in Fig 1.**

(b) Now let's linearize $I_C$ in the neighborhood of an input voltage $V_{\text{in}}^\star$ and a specific $I_C^\star$. Assume that you have a found a particular pair of input voltage $V_{\text{in}}^\star$ and current $I_C^\star$ that satisfy the current equation (9).

We can look at nearby input voltages and see how much the current changes. We can write the linearized expression for the collector current around this point as:

$$I_C(V_{\text{in}}) = I_C(V_{\text{in}}^\star) + g_m(V_{\text{in}} - V_{\text{in}}^\star) = I_C^\star + g_m\,\delta V_{\text{in}} \tag{10}$$

where $\delta V_{\text{in}} = V_{\text{in}} - V_{\text{in}}^\star$ is the change in input voltage, and $g_m$ is the slope of the local linearization around $(V_{\text{in}}^\star, I_C^\star)$. **What is $g_m$ here as a function of $I_C^\star$ and $V_{\text{TH}}$?**

*(HINT: Find $g_m$ by taking the appropriate derivative around the equilibrium point. You should recognize a part of your equation is equal to the current equilibrium point $I_C^\star = I_C(V_{\text{in}}^\star)$, so your final form should not depend on $I_S$.)*

(c) We now have a linear relationship between small changes in current and voltage, $\delta I_C = g_m\,\delta V_{\text{in}}$ around a known solution $(V_{\text{in}}^\star, I_C^\star)$.

As a reminder, the goal of this problem is to pick a particular point $(V_{\text{in}}^\star, V_{\text{out}}^\star)$ so that any small variation $\delta V_{\text{in}}$ in the input voltage $V_{\text{in}}$ can be amplified to a relatively larger variation $\delta V_{\text{out}}$ in the output voltage $V_{\text{out}}$. In other words, if $V_{\text{in}} = V_{\text{in}}^\star + \delta V_{\text{in}}$ and $V_{\text{out}} = V_{\text{out}}^\star + \delta V_{\text{out}}$, then we want the magnitude of the "amplification gain" given by $\left|\frac{\delta V_{\text{out}}}{\delta V_{\text{in}}}\right|$ to be large.

Plug your linearized equation for $I_C$ into the answer from part (a). It may help to define the output voltage equilibrium point as $V_{\text{out}}^\star$, where

$$V_{\text{out}}^\star = V_{\text{DD}} - RI_C^\star \tag{11}$$

so that we can view $V_{\text{out}} = V_{\text{out}}^\star + \delta V_{\text{out}}$ when we have $V_{\text{in}} = V_{\text{in}}^\star + \delta V_{\text{in}}$.

**Find the linearized relationship between $\delta V_{\text{out}}$ and $\delta V_{\text{in}}$.** The ratio $\frac{\delta V_{\text{out}}}{\delta V_{\text{in}}}$ is called the "small-signal voltage gain" of this amplifier around this equilibrium point.

(d) Assuming that $V_{\text{DD}} = 10\,\text{V}$, $R = 1\,\text{k}\Omega$, and $I_C^\star = 1\,\text{mA}$ when $V_{\text{in}}^\star = 0.65\,\text{V}$, **verify that the magnitude of the small-signal voltage gain $\left|\frac{\delta V_{\text{out}}}{\delta V_{\text{in}}}\right|$ is approximately** 38**.**

Next, if $I_C^\star = 9\,\text{mA}$ when $V_{\text{in}}^\star = 0.7\,\text{V}$ with all other parameters remaining fixed, **verify that the magnitude of the small-signal voltage gain $\left|\frac{\delta V_{\text{out}}}{\delta V_{\text{in}}}\right|$ between the input and the output around this equilibrium point is approximately** 346**.**

*(HINT: Remember $V_{\text{TH}} = 26\,\text{mV}$.)*

(e) If you wished to make an amplifier with as large of a small signal gain as possible, **which operating (bias) point would you choose among $V_{\text{in}}^\star = 0.65\,\text{V}$ and $V_{\text{in}}^\star = 0.7\,\text{V}$?**

This shows you that by appropriately biasing (choosing an equilibrium point), we can adjust our gain for small signals. While we just wanted to show you a simple application of linearization here, these ideas are developed a lot further in EE105, EE140, and other courses to create things like op-amps and other analog information-processing systems. Simple voltage amplifier circuits like these are used in everyday circuits like the sensors in your smartwatch, wireless transceivers in your phone, and communication circuits in CPUs and GPUs.

**Contributors:**

- Nikhil Shinde.
- Nathan Lambert.
- Anant Sahai.
- Daniel Abraham.
- Wahid Rahman.
- Sally Hui.
- Anish Muthali.
- Kris Pister.
- Alex Devonport.
- Regina Eckert.