Project Report
Application of Linear Algebra:

Least Squares Approximation Using QR decomposition.

Contents

1 INTRODUCTION

1.1 Least Squares Problem

The Least Squares problem is one of the central problems in linear algebra. This problem is best understood as a method to solve statistical problems involving linear regression, best fit and approximation.

Assume we want to find solution for a linear system of equations as below

$$\begin{cases} A_{11}x_1 & + & A_{12}x_2 & + & \dots & + & A_{1n}x_n & = & b_1 \\ A_{21}x_1 & + & A_{22}x_2 & + & \dots & + & A_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ A_{m1}x_1 & + & A_{m2}x_2 & + & \dots & + & A_{mn}x_n & = & b_m \end{cases} \Leftrightarrow \begin{cases} \sum_{j=1}^{n} A_{1j}x_j - b_1 = 0 \\ \sum_{j=1}^{n} A_{2j}x_j - b_2 = 0 \\ \vdots \\ \sum_{j=1}^{n} A_{mj}x_j - b_m = 0 \end{cases} \tag{1.1}$$

where A_{ij} and b_{ij} for $i = 1, \dots, m$ and $j = 1, \dots, n$ are known parameters, x_j for $j = 1, \dots, n$ are variables we need to find.

We rewrite (1.1) in the form of matrix, then

$$\underbrace{\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{b} \tag{1.2}$$

$$\Leftrightarrow Ax = b \tag{1.3}$$

where A is an $m \times n$ matrix, x is an $n \times 1$ matrix, b is an $m \times 1$ matrix.

If $m > n$ (normally $m >> n$ in real life application since we have more data points than the interested variables) and $b \neq 0$. Therefore, there will be more equations than the variables.

In general, we can never expect equality $Ax = b$ to hold if $m > n$. We can only expect to find a solution $x = \hat{x}$ such that $A\hat{x} \approx b$. Hence, we want to find vector x that minimizes the norm of vector $(Ax - b)$ or equivalently minimizes

$$\|Ax - b\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j - b_i \right)^2 \tag{1.4}$$

This is where the name “Least squares” comes from.

We have some methods to find the solution of the least squares problem (For this solution to be unique, the matrix A needs to have full column rank).

1.2 Using QR decomposition [1, 2, 3]

One means to solve Least-Squares problems is using QR decomposition.

For any matrix A, we can decompose A into two matrix Q and R, which means

$$\underbrace{A}_{m\times n} = \underbrace{Q}_{m\times m}\underbrace{R}_{m\times n} = \begin{pmatrix} \underbrace{Q_1}_{m\times n} & \underbrace{Q_2}_{(m-n)\times n} \end{pmatrix} \underbrace{\begin{pmatrix} R_1 \\ 0 \end{pmatrix}}_{m\times n} = \underbrace{Q_1}_{m\times n}\underbrace{R_1}_{n\times n} \tag{1.5}$$

where Q is an $m \times m$ unitary or orthogonal matrix ($Q^T = Q^{-1}$), Q_1 is an $m \times n$ matrix, Q_2 is an $(m-n) \times n$ matrix, Q_1 and Q_2 both have orthogonal columns, R_1 is an $n \times n$ upper triangular matrix and 0 is an $(m-n) \times n$ zero matrix. Note that no matter the structure of A, R_1 will always be square.

Q_2 is not unique in general. If A is of full rank n and we require that the diagonal elements of R_1 are positive then Q_1 and R_1 are unique.

QR is called full QR decomposition of matrix A and Q_1R_1 is called thin or reduced QR decomposition of A.

(a) Using Gram-Schmidt Process

Consider matrix A with n column vectors such that:

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \tag{1.6}$$

The Gram-Schmidt process proceeds by finding the orthogonal projection of the first column vector a_1

$$v_1 = a_1, \qquad e_1 = \frac{v_1}{\|v_1\|} \tag{1.7}$$

Because a_1 is the first column vector, there is no preceding projections to substract. The second column a_2 is substracted by the previous projection on the column vector:

$$v_2 = a_2 - proj_{v_1}(a_2) = a_2 - (a_2 \cdot e_1)\, e_1, \qquad e_2 = \frac{v_2}{\|v_2\|} \tag{1.8}$$

This process continues up to the n column vectors, where each incremental step $k+1$ is computed as

$$v_{k+1} = a_{k+1} - (a_{k+1} \cdot e_1)\, e_1 - \cdots - (a_{k+1} \cdot e_k)\, e_k, \qquad e_{k+1} = \frac{v_{k+1}}{\|v_{k+1}\|} \tag{1.9}$$

Thus matrix A can be factorized into the QR matrix as the following

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} = \begin{pmatrix} e_1 & e_2 & \dots & e_n \end{pmatrix} \begin{pmatrix} a_1 \cdot e_1 & a_2 \cdot e_1 & \dots & a_n \cdot e_1 \\ 0 & a_2 \cdot e_2 & \dots & a_n \cdot e_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_n \cdot e_n \end{pmatrix} = Q_1 R_1 \tag{1.10}$$

Note: By using Gram-Schmidt process, we will obtain a reduced or thin QR decomposition of A as defined in equation (1.5).

(b) Using Householder Reflections

The goal is to introduce zero under each leading entry in a column using a linear transformation. The transformation describes a reflection about a plane or hyperplane containing the origin.

Let x be an arbitrary real $m-$dimensional column vector of A. I is an $m \times m$ identity matrix, e_1 is the vector $\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}^T$. We introduce zero by changing a vector x into a vector of the same length which is collinear with e1(a standard unit vector). Set

$$u = x - sign(x)\|x\|e_1, \qquad v = \frac{u}{\|u\|} \tag{1.11}$$

where $sign(x)$ should get the opposite sign as the k-th coordinate of x, where x_k is to be the pivot coordinate after which all entries are 0 in the final upper triangular form of matrix A.

$$H = I - 2vv^T \tag{1.12}$$

If we do not want to normalize vector u, we can use the following formula

$$H = I - 2\frac{uu^T}{u^T u} \tag{1.13}$$

where H is an $m \times m$ Householder matrix, which is both symmetric and orthogonal and

$$Hx = \begin{pmatrix} \|x\| \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1.14}$$

This can be used to gradually transform an $m \times n$ matrix A to upper triangular form. First, we multiply A with the Householder matrix H_1 we obtain when we choose first matrix column for x. This results in a matrix $H_1 A$ with zeros in the left column

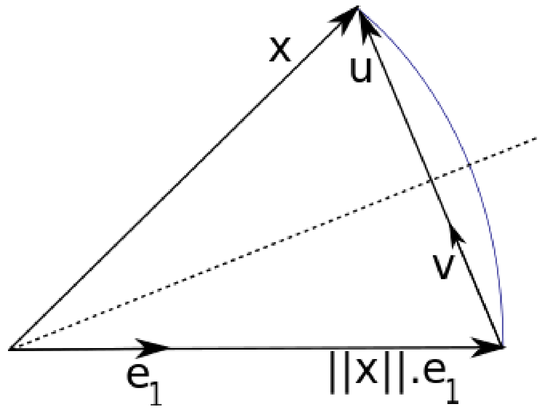

Figure 1: Householder reflection method for QR decomposition

(except for the first row).

$$H_1 A = \begin{pmatrix} \|x_1\| & \times & \dots & \times \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix} \tag{1.15}$$

This can be repeated for A' (obtained from H_1A by deleting the first row and first column), resulting in a Householder matrix H'_2. Note that H'_2 is smaller than H_1. Since we want it to operate on H_1A instead of A' we need to expand it to the upper left, filling with identity matrix, in general

$$H_k = \begin{pmatrix} I_{k-1} & 0 \\ 0 & H'_k \end{pmatrix} \tag{1.16}$$

After t iterations of this process, $t = min(m-1, n)$,

$$R = H_t \dots H_2 H_1 A \tag{1.17}$$

is an upper triangular matrix, and let

$$Q = H_1^T H_2^T \dots H_t^T = H_1 H_2 \dots H_t \tag{1.18}$$

Hence, $A = QR$ is a QR decomposition of A.

→ This method has greater numerical stability than the Gram-Schmidt method.

Note: By using Householder reflection method, we will obtain a full QR decomposition of A as defined in equation (1.5).

(c) Using Givens Rotations

A Givens rotation is presented by a matrix of the form:

$$G(i,j,\theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & -s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \tag{1.19}$$

where $c = \cos\theta$ and $s = \sin\theta$ appear at the intersections ith and jth rows and columns. That is, for fixed $i > j$, the non-zero elements of Givens matrix are given by

$$\begin{aligned} g_{kk} &= 1 \qquad \text{for } k \neq i,\ j \\ g_{kk} &= c \qquad \text{for } k = i,\ j \\ g_{ji} &= -g_{ij} = -s \end{aligned} \tag{1.20}$$

The product $G(i,j,\theta)x$ represents a counterclockwise rotation of the vector x in the (i,j) plane of θ radians, hence the name Givens rotation. $G(i,j)$ is orthogonal.

When a Givens rotation matrix, $G(i,j,\theta)$, multiplies another matrix, A, from the left, GA, only rows i and j of A are affected. Thus we only pay attention to the following counter-clockwise problem. Given a and b, find c and s such that

$$\begin{pmatrix} c & -s \\ s & c \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} r \\ 0 \end{pmatrix} \tag{1.21}$$

where $r = \sqrt{a^2 + b^2}$, we can find an obvious solution: $c = \frac{a}{r}$, and $s = -\frac{b}{r}$

Method: Introduce zero in each column (columns from left to right) from the bottom entry to the entry below the leading entry by a matrix G.

Suppose we want to introduce zero in entry (i,j) in A, then G is a $m \times m$ matrix with $\begin{pmatrix} c & -s \\ s & c \end{pmatrix}$ lies on the diagonal line on row$(i-1)$ and row(i). All other entries in diagonal lines are 1 and the rest of entries are zero.

After t iterations of rotations to transform A into upper triangular matrix R, we can obtain:

$$R = G_t \ldots G_2 G_1 A \tag{1.22}$$

and

$$Q = G_1^T G_2^T ... G_t^T \tag{1.23}$$

Note: By using Givens rotation method, we will obtain a full QR decomposition of A as defined in equation (1.5).

Least Squares solution using QR decomposition:

Use the fact that multiplication by orthogonal matrices does not change Euclidean lengths, since Q^T is orthogonal matrix, we have

$$\|Ax-b\|^2 = \left\|Q^T(Ax-b)\right\|^2 \tag{1.24}$$

$$= \left\| \begin{pmatrix} R_1 \\ 0 \end{pmatrix} x - \begin{pmatrix} Q_1^T b \\ Q_2^T b \end{pmatrix} \right\|^2 \tag{1.25}$$

$$= \left\|R_1 x - Q_1^T b\right\|^2 + \left\|Q_2^T b\right\|^2 \tag{1.26}$$

$\|Ax-b\|^2$ is minimum when $\left\|R_1 x - Q_1^T b\right\|^2 = 0$ or $R_1 x - Q_1^T b = 0$

The solution for Least Squares problem using QR decomposition could be expressed as

$$\boxed{\hat{x} = R_1^{-1}(Q_1^T b)} \tag{1.27}$$

Note: In practice, A is a very long matrix ($n \gg m$), probability that A is full rank is very likely. Therefore, R_1 is unique and we usually obtain a unique solution for a given real life problem.

2 SMALL EXAMPLE

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 2 & 3 & 1 \\ 4 & 1 & 2 \end{pmatrix} \tag{2.1}$$

We will factorise matrix A above using three methods of QR decomposition.

1. **Gram-Schmidt method:**

$$v_1 = a_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix}, \qquad e_1 = \frac{v_1}{\|v_1\|} = \frac{1}{5.4772} \begin{pmatrix} 1 \\ 3 \\ 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 0.1826 \\ 0.5477 \\ 0.3651 \\ 0.7303 \end{pmatrix} \tag{2.2}$$

$$\rightarrow a_1 \cdot e_1 = a_1^T e_1 = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix} \begin{pmatrix} 0.1826 \\ 0.5477 \\ 0.3651 \\ 0.7303 \end{pmatrix} = 4.3818 \tag{2.3}$$

$$v_2 = a_2 - (a_2 \cdot e_2) e_2 = \begin{pmatrix} 2 \\ 4 \\ 3 \\ 1 \end{pmatrix} - 4.3818 \times \begin{pmatrix} 0.1826 \\ 0.5477 \\ 0.3651 \\ 0.7303 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 6 \\ 8 \\ 7 \\ -11 \end{pmatrix} \tag{2.4}$$

$$\rightarrow e_2 = \frac{v_2}{\|v_2\|} = \frac{1}{3.2863} \times \frac{1}{5} \begin{pmatrix} 6 \\ 8 \\ 7 \\ -11 \end{pmatrix} = \begin{pmatrix} 0.3651 \\ 0.4869 \\ 0.4260 \\ -0.6694 \end{pmatrix} \tag{2.5}$$

$$\rightarrow a_3 \cdot e_1 = a_3^T e_1 = \begin{pmatrix} 3 & 2 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0.1826 \\ 0.5477 \\ 0.3651 \\ 0.7303 \end{pmatrix} = 3.4689 \tag{2.6}$$

$$\rightarrow a_3 \cdot e_2 = a_3^T e_2 = \begin{pmatrix} 3 & 2 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0.3651 \\ 0.4869 \\ 0.4260 \\ -0.6694 \end{pmatrix} = 1.1563 \tag{2.7}$$

$$v_3 = a_3 - (a_3 \cdot e_1)e_1 = (a_3 \cdot e_2)e_2 = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix} - 3.4689 \begin{pmatrix} 0.1826 \\ 0.5477 \\ 0.3651 \\ 0.7303 \end{pmatrix} - 1.1563 \begin{pmatrix} 0.3651 \\ 0.4869 \\ 0.4260 \\ -0.6694 \end{pmatrix} \tag{2.8}$$

$$\rightarrow v_3 = \frac{1}{54} \begin{pmatrix} 105 \\ -25 \\ -41 \\ 13 \end{pmatrix}, \qquad e_3 = \frac{v_3}{\|v_3\|} = \frac{1}{2.1517} \times \frac{1}{54} \begin{pmatrix} 105 \\ -25 \\ -41 \\ 13 \end{pmatrix} = \begin{pmatrix} 0.9037 \\ -0.2152 \\ -0.3529 \\ 0.1119 \end{pmatrix} \tag{2.9}$$

Hence,

$$Q_1 = \begin{pmatrix} e_1 & e_2 & e_3 \end{pmatrix} = \begin{pmatrix} 0.1826 & 0.3651 & 0.9037 \\ 0.5477 & 0.4869 & -0.2152 \\ 0.3651 & 0.4260 & -0.3529 \\ 0.7303 & -0.6694 & 0.1119 \end{pmatrix} \tag{2.10}$$

and

$$R_1 = \begin{pmatrix} a_1 \cdot e_1 & a_2 \cdot e_1 & a_3 \cdot e_1 \\ 0 & a_2 \cdot e_2 & a_3 \cdot e_2 \\ 0 & 0 & a_3 \cdot e_3 \end{pmatrix} = \begin{pmatrix} \|v_1\| & a_2 \cdot e_1 & a_3 \cdot e_1 \\ 0 & \|v_2\| & a_3 \cdot e_2 \\ 0 & 0 & \|v_3\| \end{pmatrix} \tag{2.11}$$

$$\rightarrow R_1 = \begin{pmatrix} 5.4773 & 4.3818 & 3.4689 \\ 0 & 3.2863 & 1.1563 \\ 0 & 0 & 2.1517 \end{pmatrix} \tag{2.12}$$

Note: We can observe that Gram-Schmidt process will give $A = Q_1 R_1$ which is a reduced QR decomposition of A (Q_1 is orthogonal matrix, R_1 is a square upper triangular matrix as defined in equation (1.5)).

2. **Householder Reflection method:**

First, let $x_1 = \begin{pmatrix} 1 & 3 & 2 & 4 \end{pmatrix}^T$ be the first column vector of matrix A, then we obtain

$$A_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 2 & 3 & 1 \\ 4 & 1 & 2 \end{pmatrix} \tag{2.13}$$

$$\|x_1\| = \sqrt{1^2 + 3^2 + 2^2 + 4^2} = \sqrt{30} = 5.4772 \tag{2.14}$$

$$\rightarrow u_1 = x_1 + \|x_1\| \, e_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + 5.4772 \times \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6.4772 \\ 3 \\ 2 \\ 4 \end{pmatrix} \tag{2.15}$$

$$\rightarrow H_1 = I - 2\frac{u_1 u_1^T}{u_1^T u_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \frac{2}{70.9545} \times \begin{pmatrix} 6.4772 \\ 3 \\ 2 \\ 4 \end{pmatrix} \cdot \begin{pmatrix} 6.4772 & 3 & 2 & 4 \end{pmatrix} \tag{2.16}$$

$$\rightarrow H_1 = H_1' = \begin{pmatrix} -0.1826 & -0.5477 & -0.3651 & -0.7303 \\ -0.5477 & 0.7463 & -0.1691 & -0.3382 \\ -0.3651 & -0.1691 & 0.8873 & -0.2255 \\ -0.7303 & -0.3382 & -0.2255 & 0.5490 \end{pmatrix} \tag{2.17}$$

$$H_1 A_1 = H_1' A_1 = \begin{pmatrix} -0.1826 & -0.5477 & -0.3651 & -0.7303 \\ -0.5477 & 0.7463 & -0.1691 & -0.3382 \\ -0.3651 & -0.1691 & 0.8873 & -0.2255 \\ -0.7303 & -0.3382 & -0.2255 & 0.5490 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 2 & 3 & 1 \\ 4 & 1 & 2 \end{pmatrix} \tag{2.18}$$

$$\rightarrow H_1 A_1 = \begin{pmatrix} -5.4772 & -4.3818 & -3.4689 \\ 0 & 1.0442 & -0.9961 \\ 0 & 1.0295 & -1.9949 \\ 0 & -2.9411 & -1.9949 \end{pmatrix} \tag{2.19}$$

Now removing first row and first column, we get

$$A_2 = \begin{pmatrix} 1.0442 & -0.9961 \\ 1.0295 & -1.9949 \\ -2.9411 & -1.9949 \end{pmatrix} \tag{2.20}$$

Thus, let $x_2 = \begin{pmatrix} 1.0442 & 1.0295 & -2.9411 \end{pmatrix}^T$, then

$$\|x_2\| = \sqrt{1.0442^2 + 1.0295^2 + 2.9411^2} = \sqrt{\frac{54}{5}} = 3.2863 \tag{2.21}$$

$$\rightarrow u_2 = x_2 + \|x_2\|\, e_1 = \begin{pmatrix} 1.0442 \\ 1.0295 \\ -2.9411 \end{pmatrix} + 3.2863 \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4.3305 \\ 1.0295 \\ -2.9411 \end{pmatrix} \tag{2.22}$$

$$H_2' = I - 2\frac{v_2 v_2^T}{v_2^T v_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{2}{28.4632} \begin{pmatrix} 4.3305 \\ 1.0295 \\ -2.9411 \end{pmatrix} \cdot \begin{pmatrix} 4.3305 & 1.0295 & -2.9411 \end{pmatrix} \tag{2.23}$$

$$\rightarrow H_2' = \begin{pmatrix} -0.3177 & -0.3133 & 0.8949 \\ -0.3133 & 0.9255 & 0.2127 \\ 0.8949 & 0.2128 & 0.3992 \end{pmatrix} \tag{2.24}$$

$$H_2' A_2 = \begin{pmatrix} -0.3177 & -0.3133 & 0.8949 \\ -0.3133 & 0.9255 & 0.2127 \\ 0.8949 & 0.2128 & 0.3992 \end{pmatrix} \cdot \begin{pmatrix} 1.0442 & -0.9961 \\ 1.0295 & -1.9949 \\ -2.9411 & -1.9949 \end{pmatrix} \tag{2.25}$$

$$\rightarrow H_2' A_2 = \begin{pmatrix} -3.2863 & -1.1563 \\ 0 & -1.0355 \\ 0 & -1.8861 \end{pmatrix} \tag{2.26}$$

Now removing 1st row and 1st column, we obtain

$$A_3 = \begin{pmatrix} -1.0355 \\ -1.8861 \end{pmatrix} \tag{2.27}$$

$$\rightarrow x_3 = \begin{pmatrix} -1.0355 \\ -1.8861 \end{pmatrix}, \qquad \|x_3\| = \sqrt{1.0355^2 + 1.8861^2} = \sqrt{\frac{125}{27}} = 2.1517 \tag{2.28}$$

$$u_3 = x_3 - \|x_3\|e_1 = \begin{pmatrix} -1.0355 \\ -1.8861 \end{pmatrix} - 2.1517 \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -3.1872 \\ -1.8861 \end{pmatrix} \tag{2.29}$$

$$H_3' = I - 2\frac{u_3 u_3^T}{v_3^T v_3} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2}{13.7154} \begin{pmatrix} -3.1872 \\ -1.8861 \end{pmatrix} \cdot \begin{pmatrix} -3.1872 & -1.8861 \end{pmatrix} \tag{2.30}$$

$$\to H_3' = \begin{pmatrix} -0.4813 & -0.8766 \\ -0.8766 & 0.4813 \end{pmatrix} \tag{2.31}$$

$$H_3' A_3 = \begin{pmatrix} -0.4813 & -0.8766 \\ -0.8766 & 0.4813 \end{pmatrix} \cdot \begin{pmatrix} -1.0355 \\ -1.8861 \end{pmatrix} = \begin{pmatrix} 2.1517 \\ 0 \end{pmatrix} \tag{2.32}$$

Since $H_3 H_2 H_1 A = R$,

$$\begin{aligned} R = & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.4813 & -0.8766 \\ 0 & 0 & -0.8766 & 0.4813 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.3177 & -0.3133 & 0.8949 \\ 0 & -0.3133 & 0.9255 & 0.2127 \\ 0 & 0.8949 & 0.2128 & 0.3992 \end{pmatrix} \\ & \times \begin{pmatrix} -0.1826 & -0.5477 & -0.3651 & -0.7303 \\ -0.5477 & 0.7463 & -0.1691 & -0.3382 \\ -0.3651 & -0.1691 & 0.8873 & -0.2255 \\ -0.7303 & -0.3382 & -0.2255 & 0.5490 \end{pmatrix} \times \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 2 & 3 & 1 \\ 4 & 1 & 2 \end{pmatrix} \end{aligned} \tag{2.33}$$

$$\to R = \begin{pmatrix} -5.4772 & -4.3818 & -3.4689 \\ 0 & -3.2863 & -1.1563 \\ 0 & 0 & 2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.34}$$

$$Q = H_1H_2H_3 = \begin{pmatrix} -0.1826 & -0.5477 & -0.3651 & -0.7303 \\ -0.5477 & 0.7463 & -0.1691 & -0.3382 \\ -0.3651 & -0.1691 & 0.8873 & -0.2255 \\ -0.7303 & -0.3382 & -0.2255 & 0.5490 \end{pmatrix}$$

$$\times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.3177 & -0.3133 & 0.8949 \\ 0 & -0.3133 & 0.9255 & 0.2127 \\ 0 & 0.8949 & 0.2128 & 0.3992 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.4813 & -0.8766 \\ 0 & 0 & -0.8766 & 0.4813 \end{pmatrix} \tag{2.35}$$

$$\to Q = \begin{pmatrix} -0.1826 & -0.3651 & 0.9037 & -0.1219 \\ -0.5477 & -0.4869 & -0.2152 & 0.6455 \\ -0.3651 & -0.4260 & -0.3529 & -0.7488 \\ -0.7303 & 0.6694 & 0.1119 & -0.0775 \end{pmatrix} \tag{2.36}$$

Note: We can observe that Householder reflection method will give $A = QR$ which is a full QR decomposition of A (Q has orthogonal columns, R is an upper triangular matrix as defined in equation (1.5)).

3. **Givens Rotation method:**

Step 1: Introduce zero in entry $(4, 1)$ of matrix A.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \to r = \sqrt{a^2 + b^2} = \sqrt{20},\ c = \frac{a}{r} = \frac{\sqrt{5}}{5},\ s = -\frac{b}{r} = -\frac{2\sqrt{5}}{5} \tag{2.37}$$

$$\to G_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.4472 & 0.8944 \\ 0 & 0 & -0.8944 & 0.4472 \end{pmatrix} \tag{2.38}$$

$$\to R_1' = G_1 A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 4 & 2 \\ 4.4721 & 2.2361 & 2.2361 \\ 0 & -2.2362 & 0 \end{pmatrix} \tag{2.39}$$

Step 2: Introduce zero in entry $(3,1)$ of matrix R_1'.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3 \\ 4.4721 \end{pmatrix} \to c = 0.5571,\ s = -0.8305 \tag{2.40}$$

$$\to G_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5571 & 0.8305 & 0 \\ 0 & -0.8305 & 0.5571 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.41}$$

$$\to R_2' = G_2 R_1' = \begin{pmatrix} 1 & 2 & 3 \\ 5.2853 & 4.0854 & 2.9712 \\ 0 & -2.0735 & -0.4153 \\ 0 & -2.2361 & 0 \end{pmatrix} \tag{2.42}$$

Step 3: Introduce zero in entry $(2,1)$ of matrix R_2'.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 5.3853 \end{pmatrix} \to c = 0.1826,\ s = -0.9832 \tag{2.43}$$

$$\to G_3 = \begin{pmatrix} 0.1826 & 0.9832 & 0 & 0 \\ -0.9832 & 0.1826 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.44}$$

$$\to R_3' = G_3 R_2' = \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & -1.2204 & -2.4071 \\ 0 & -2.0735 & -0.4153 \\ 0 & -2.2361 & 0 \end{pmatrix} \tag{2.45}$$

Step 4: Introduce zero in entry $(4,2)$ of matrix R_3'.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -2.0735 \\ -2.2361 \end{pmatrix} \to c = -0.6804,\ s = 0.7328 \tag{2.46}$$

$$\rightarrow G_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.6804 & -0.7328 \\ 0 & 0 & 0.7328 & -0.6804 \end{pmatrix} \tag{2.47}$$

$$\rightarrow R_4' = G_4 R_3' = \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & -1.2204 & -2.4071 \\ 0 & 3.0513 & 0.2825 \\ 0 & 0 & -0.3043 \end{pmatrix} \tag{2.48}$$

Step 5: Introduce zero in entry $(3, 2)$ of matrix R_4'.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -1.2204 \\ 2.0513 \end{pmatrix} \rightarrow c = -0.3714,\ s = -0.9285 \tag{2.49}$$

$$\rightarrow G_5 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.3714 & 0.9285 & 0 \\ 0 & -0.9285 & -0.3714 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.50}$$

$$\rightarrow R_5' = G_5 R_4' = \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & 3.2863 & 1.1562 \\ 0 & 0 & 2.1301 \\ 0 & 0 & -0.3043 \end{pmatrix} \tag{2.51}$$

Step 6: Introduce zero in entry $(4, 3)$ of matrix R_5'.

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2.1301 \\ -0.3043 \end{pmatrix} \rightarrow c = 0.9899,\ s = 0.1414 \tag{2.52}$$

$$\rightarrow G_6 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.9899 & -0.1414 \\ 0 & 0 & 0.1414 & 0.9899 \end{pmatrix} \tag{2.53}$$

$$\rightarrow R'_6 = G_6 R'_5 = \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & 3.2863 & 1.1562 \\ 0 & 0 & 2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.54}$$

Hence,

$$R = R'_6 = G_6 G_5 G_4 G_3 G_2 G_1 A = \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & 3.2863 & 1.1562 \\ 0 & 0 & 2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.55}$$

and

$$Q = G_1^T G_2^T G_3^T G_4^T G_5^T G_6^T = \begin{pmatrix} 0.1826 & 0.3652 & 0.9037 & 0.8021 \\ 0.5477 & 0.4869 & -0.2152 & -0.6455 \\ 0.3652 & 0.4260 & -0.3528 & 0.7587 \\ 0.7303 & -0.6694 & 0.1119 & 0.0774 \end{pmatrix} \tag{2.56}$$

$\rightarrow$ **Evaluation:**

From equations (2.10), (2.12), (2.34), (2.36), we can observe that using Gram-Schmidt or Householder reflections to decompose matrix A will give the same result. Specifically,

Using Gram-Schmidt process, we obtain:

$$A = Q_1 R_1 = \begin{pmatrix} 0.1826 & 0.3651 & 0.9037 \\ 0.5477 & 0.4869 & -0.2152 \\ 0.3651 & 0.4260 & -0.3529 \\ 0.7303 & -0.6694 & 0.1119 \end{pmatrix} \begin{pmatrix} 5.4773 & 4.3818 & 3.4689 \\ 0 & 3.2836 & 1.1563 \\ 0 & 0 & 2.1517 \end{pmatrix} \tag{2.57}$$

Using Householder reflection method, we obtain

$$A = QR = \begin{pmatrix} -0.1826 & -0.3651 & 0.9037 & -0.1219 \\ -0.5477 & -0.4869 & -0.2152 & 0.6455 \\ -0.3651 & -0.4260 & -0.3529 & -0.7488 \\ -0.7303 & 0.6694 & 0.1119 & -0.0775 \end{pmatrix} \begin{pmatrix} -5.4772 & -4.3818 & -3.4689 \\ 0 & -3.2863 & -1.1563 \\ 0 & 0 & 2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.58}$$

$$= \begin{pmatrix} 0.1826 & 0.3651 & 0.9037 & -0.1219 \\ 0.5477 & 0.4869 & -0.2152 & 0.6455 \\ 0.3651 & 0.4260 & -0.3529 & -0.7488 \\ 0.7303 & -0.6694 & 0.1119 & -0.0775 \end{pmatrix} \begin{pmatrix} +5.4772 & 4.3818 & 3.4689 \\ 0 & +3.2863 & 1.1563 \\ 0 & 0 & +2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.59}$$

$$= \begin{pmatrix} 0.1826 & 0.3651 & 0.9037 \\ 0.5477 & 0.4869 & -0.2152 \\ 0.3651 & 0.4260 & -0.3529 \\ 0.7303 & -0.6694 & 0.1119 \end{pmatrix} \begin{pmatrix} 5.4773 & 4.3818 & 3.4689 \\ 0 & 3.2863 & 1.1563 \\ 0 & 0 & 2.1517 \end{pmatrix} = Q_1 R_1 \tag{2.60}$$

Moreover, from equations (2.55) and (2.56), we get QR decomposition of A using Givens rotation method:

$$A = QR = \begin{pmatrix} 0.1826 & 0.3652 & 0.9037 & 0.8021 \\ 0.5477 & 0.4869 & -0.2152 & -0.6455 \\ 0.3652 & 0.4260 & -0.3528 & 0.7587 \\ 0.7303 & -0.6694 & 0.1119 & 0.0774 \end{pmatrix} \begin{pmatrix} 5.4774 & 4.3819 & 3.4691 \\ 0 & 3.2863 & 1.1562 \\ 0 & 0 & 2.1517 \\ 0 & 0 & 0 \end{pmatrix} \tag{2.61}$$

In general, Gram-Schmidt process will give us reduced QR decomposition while Householder reflection and Givens rotation method will give full QR decomposition as defined in equation (1.5).

We can observe that Q is not unique for Householder reflection and Givens rotation methods. However, Q_1 and R_1 are unique (except for some tolerances of 0.1-0.5% during calculating and rounding process, percentage error will be negligible when using software such as matlab or python, since they could store values with very high precision).

3 IMPLEMENTATION IN REAL LIFE

In this section of our report, we wish to illustrate one of many possible applications of the QR Decomposition algorithm. More specifically, we will elaborate on how we use the QR Decomposition algorithm in order to solve the least squares approximation problem. We have provided 2 case studies of implementation of QR Decomposition algorithm in Python, Factors of Happiness, and Prices in the Housing Market in the United States.

In each of these implementations, the general flow is as follow:

1. Firstly, we import relevant modules (Figure 2):
   - Numpy: to perform matrix and linear algebra operations including QR Decomposition algorithms
   - Matplotlib.pyplot: for data visualisation (graphing)
   - Math: to perform arithmetic operations
   - Csv: to read data (which is in csv)

```
import numpy as np
import matplotlib.pyplot as plt
import math
import csv
```

Figure 2: Implementation Code Part 1

2. Next, we read the data set, storing the independent variable in the Y matrix and the dependent variables in the X matrix. Here, Y is a 5000×1 Column matrix while X is a 5000×5 matrix.

3. We assign Average Income as the X axis as we believe it to be the most important factor in the case of housing prices. (Figure 3)

```
for row in csvreader:
    # Dependent Variable: Prices
    Y.append([float(row[5])])
    # Independent Variables:
    # 1. Average Income in the Area
    # 2. Average House Age in the Area
    # 3. Average Number of Rooms in the Area
    # 4. Average Number of Bedrooms in the Area
    # 5. Population in the Area
    X.append([float(row[0]),
              float(row[1]),
              float(row[2]),
              float(row[3]),
              float(row[4])
              ])
    # Average Income in the Area is chosen as X Axis
    X_axis.append([float(row[0])])
```

Figure 3: Implementation Code Part 2

4. We then perform the QR Decomposition Algorithm (Figure 4)

– We first convert the X and Y arrays into matrix using numpy
– Then, we find the Q and R matrix using numpy's QR Decomposition algorithm
– Then, we find the the inverse of R and transpose of Q
– The predicted Y values will then be $XR^{-1}Q^TY$

```
# Convert X and Y arrays into matrices
# so that numpy matrix operations can be performed
X = np.array(X)
Y = np.array(Y)

# Perform QR Decomposition Algorithm
Q, R = np.linalg.qr(X)

R_inverse = np.linalg.inv(R) # Find Inverse of R
Q_transpose = Q.T # Find Transpose of Q
B = np.matmul(np.matmul(R_inverse, Q_transpose), Y) # B = Inverse of R * Transpose of Q * Y
Y_Predicted = np.matmul(X,B) # Y = X * B
```

Figure 4: Implementation Code Part 3

5. Then, we plot both the actual Y and predicted Y values against a chosen X axis, and add the appropriate labels, titles and colours (Figure 5)

```
plt.title("Price against Average Income\n in the United States Housing Market")
plt.xlabel("Average Income (USD)")
plt.ylabel("Price (1 Million USD)")
plt.scatter(X_axis,Y,color="red", label="Actual Price")
plt.scatter(X_axis,Y_Predicted,color="black", label="Predicted Price")
plt.legend(loc = "upper left")
plt.show()
```

Figure 5: Implementation Code Part 4

6. We then calculate the Mean Percentage Error (MPE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) and Adjusted R^2 to evaluate the accuracy of the prediction. MPE, MAE, MSE, RMSE and Adjusted R^2 are defined as follow:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{3.1}$$

where M is mean absolute percentage error, n is number of times the summation iteration happens, A_t is actual value, F_t is forecast value.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \tag{3.2}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{3.3}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \tag{3.4}$$

where $\hat{y}$ is predicted value of y, $\bar{y}$ is the mean value of y.

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{total sum of squares (SST)}} \tag{3.5}$$

$$= 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3.6}$$

$$\text{Adjusted } R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1} \tag{3.7}$$

where R^2 is sample R-squared, N is total sample size, p is number of independent variables.

The code employed to calculate the above analysis parameters is shown in figure 6.

7. We also perform the above steps using only a single independent variable in order to compare the accuracy of the results.

```python
MPE = 0.0 # Mean Percentage Error
MAE = 0.0 # Mean Absolute Error
MSE = 0.0 # Mean Squared Error
RMSE = 0.0 # Root Mean Squared Error

for i in range(0, 5000):
    MPE += abs((Y[i][0] - Y_Predicted[i][0]) / Y[i][0]) * 100
    MAE += abs(Y[i][0] - Y_Predicted[i][0])
    MSE += (Y[i][0] - Y_Predicted[i][0])**2

MPE = MPE / 5000.0
MAE = MAE / 5000.0
MSE = MSE / 5000.0
RMSE = MSE ** (0.5)

R_Squared = 0.0 # Coefficient of Determination
Adjusted_R_Squared = 0.0
Y_Mean = 0.0

for i in range(0, 5000):
    Y_Mean += Y[i][0]

Y_Mean = Y_Mean / 5000

SSR = 0.0 # Sum Squared Regression
TSS = 0.0 # Total Sum of Squares

for i in range(0, 5000):
    SSR += (Y[i][0] - Y_Predicted[i][0])**2
    TSS += (Y[i][0] - Y_Mean)**2

R_Squared = 1 - (SSR / TSS)
Adjusted_R_Squared = 1 - (1 - R_Squared**2) * (5000 - 1) / (5000 - 5 - 1)

print(MPE)
print(MAE)
print(MSE)
print(RMSE)
print(R_Squared)
print(Adjusted_R_Squared)
```

Figure 6: Implementation Code Part 5

Case Study 1 Factors of Happiness [4]

Dependent Variable: Happiness Score

Independent Variables:

- Economy (GDP Per Capita relative to the United States)
- Family Score
- Life Expectancy
- Freedom
- Government Trust
- Community Generosity
- Dystopia Residual

Upon implementing QR Decomposition to solve the Least Squares problem in order to model the actual data in this case study, we were able to generate predicted values of happiness score of a country based on the independent variables mentioned. The plot of actual and predicted happiness scores against the GDP per Capita of a country using multiple independent variables is shown in figure 7.

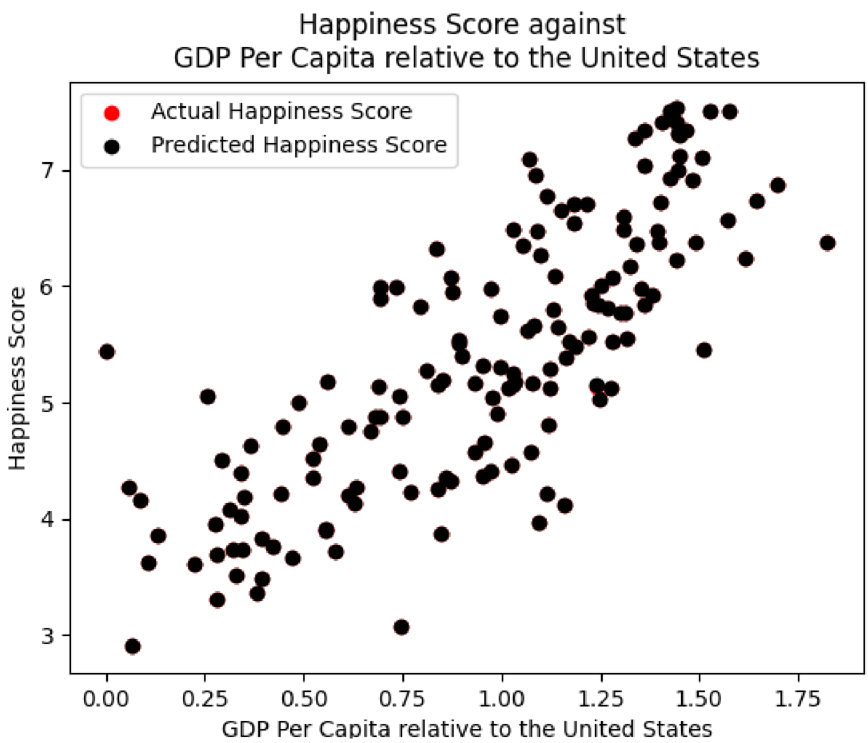

Figure 7: Plot of Actual and Predicted Happiness Score against GDP Per Capita (relative to the United States) (Multiple Independent Variables)

We also measured several analysis parameters, namely: Mean Percentage Error (MPE), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error

Analysis Parameters	Values
Mean Percentage Error (MPE)	0.004932398663796852%
Mean Average Error (MAE)	0.00025298625439121\65
Mean Squared Error (MSE)	8.658191814554322e-08
Root Mean Squared Error (RMSE)	0.00029424805546603\57
Adjusted R^2	0.99999986001332

Table 1: Analysis Parameters (Multiple Independent Variables)

(RMSE) and Adjusted R2. The values of these analysis parameters are shown in the table 1.

We also re-did the implementation while using a single independent variable, Economy (GDP Per Capita relative to the United States), in order to compare the accuracy of the results. The plot of actual and predicted happiness scores against the GDP per Capita of a country using a single independent variable is shown in figure 8. We also measured

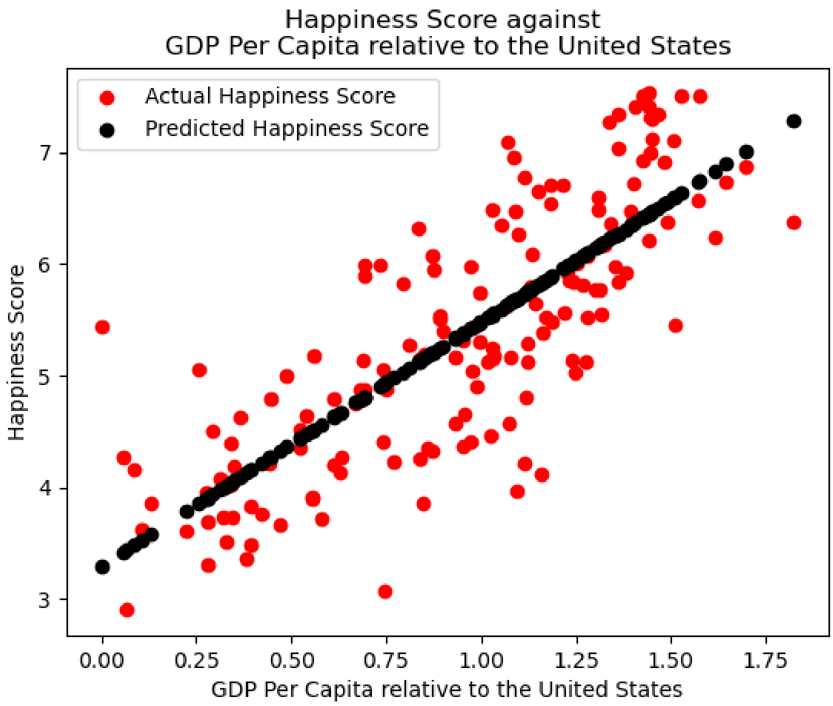

Figure 8: Plot of Actual and Predicted Happiness Score against GDP Per Capita (relative to the United States) (Single Independent Variables)

the relevant analysis parameters and recorded their values below in table 2.

→ Evaluation:

One key insight we learn from this case study is that taking into account multiple independent variables will allow us to make a more accurate models in general. This can be seen by the significantly better fit of the graphs and much lower error rates. For instance the mean percentage error using using multiple independent variables is 0.005 percent while that of using Single Independent Variables which is 23.9%.

This highlights the advantage of using QR Decomposition to solve least square problem in data modelling as compared to other methods such as single variable linear regression which only allows for one independent variable

However, we also find that adding a variable does not necessarily result in better fit. To check whether adding an independent variable result in better fit, adjusted R^2 can be used. If adding variable result in higher adjusted R^2, it means that adding that variable will result in better fit and vice versa. As can be seen here, adjusted R^2 value greatly increased from 0.47 to 0.99 upon adding the 6 relevant independent variables.

Analysis Parameters	Values
Mean Percentage Error (MPE)	23.895749714721507%
Mean Average Error (MAE)	1.1437203766602564
Mean Squared Error (MSE)	2.1893494579509056
Root Mean Squared Error (RMSE)	1.4796450445802554
Adjusted R^2	0.47336662131759466

Table 2: Analysis Parameters (Single Independent Variables)

Case Study 2: The Housing Market in the United States [5]

In this case study, the dependent and independent variables are as follow:

Dependent Variable: Housing Price

Independent Variables:

- Average Income in the Area
- Average House Age in the Area
- Average Number of Rooms in the Area
- Average Number of Bedrooms in the Area
- Population in the Area

We conducted this case study to emphasize the use of Adjusted R2 to measure whether adding an independent variable will result in better fit of predicted values. We implemented using all 5 independent variables, 4 independent variable (without Population Size) and 4 Independent variables using (without Mean Income). The plots of actual and predicted housing price against the average income of an area using 5 independent variables, 4 independent variable (without Population Size) and 4 Independent variables using (without Mean Income) are shown below in figures 9, 10 and 11 respectively.

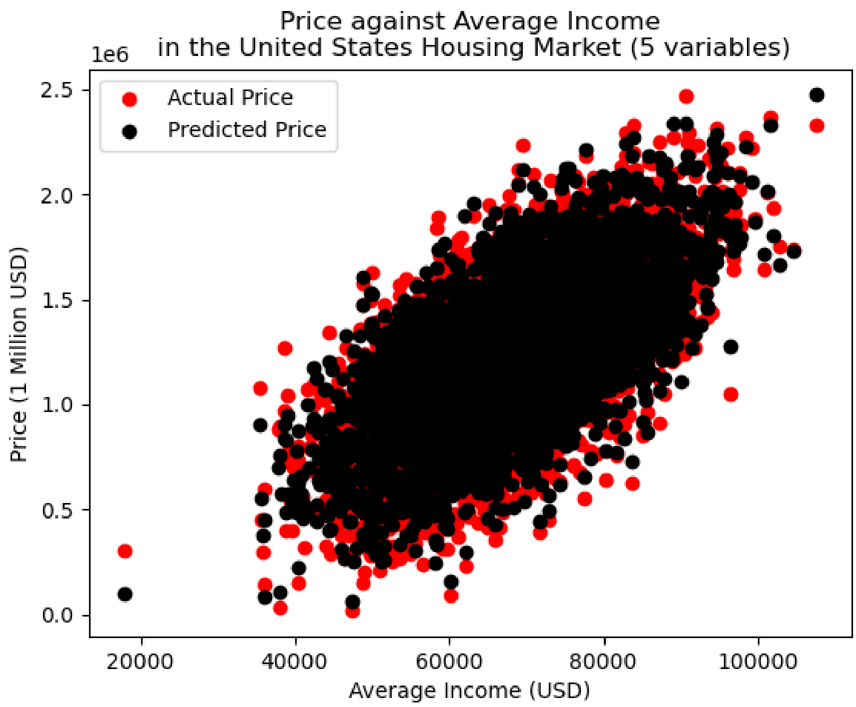

Figure 9: Plot of Actual and Predicted Housing Price against the Average Income (5 Independent Variables)

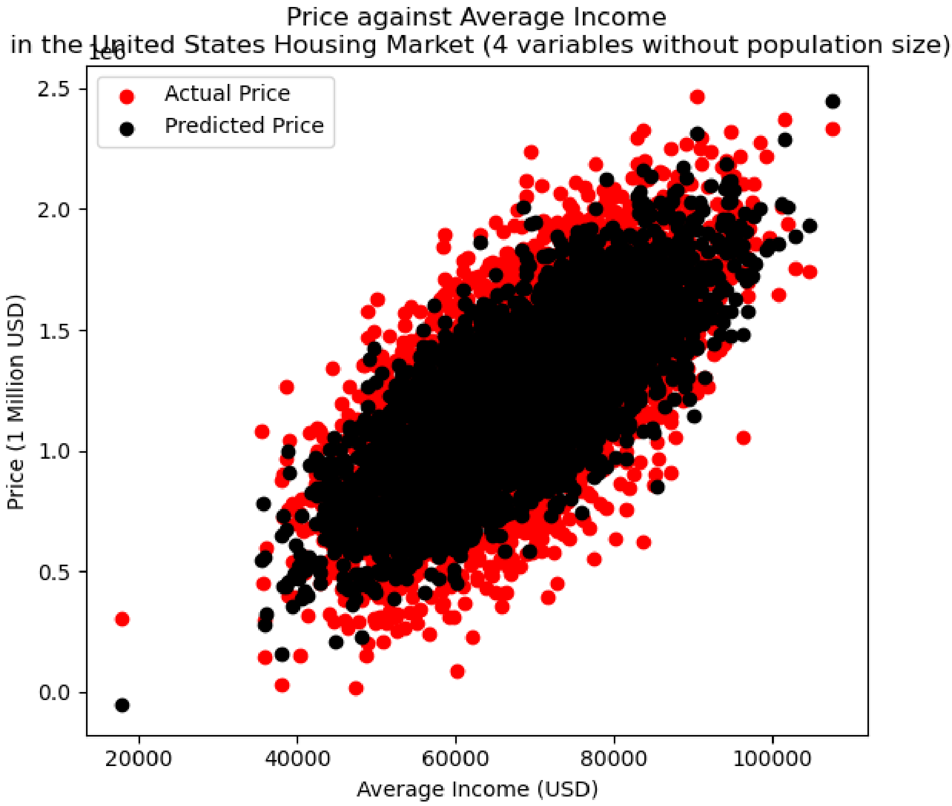

Figure 10: Plot of Actual and Predicted Housing Price against the Average Income (4 Independent Variables without Population Size)

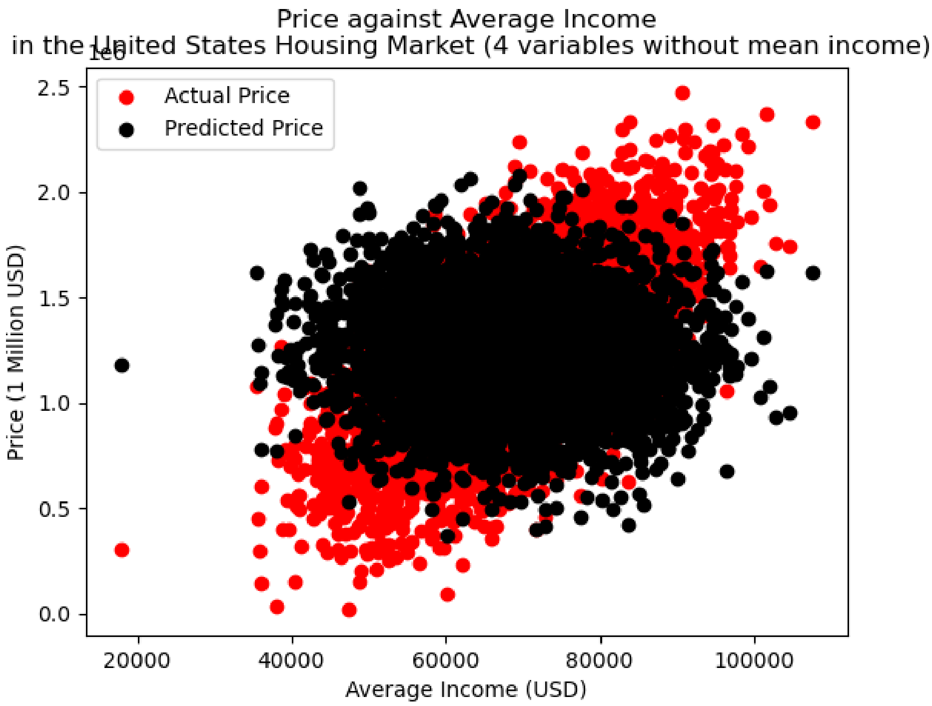

Figure 11: Plot of Actual and Predicted Housing Price against the Average Income (4 Independent Variables without Mean Income)

We provided the Mean Percentage Error (MPE) and Adjusted R2 for each of these 3 cases in table 3.

Number of varibles	MPE	Adjusted R^2
5 Independent Variables	21.421887173973253%	0.28039252840049145
4 Independent Variables (without Population Size)	22.619540352644073%	0.22273862850632908
4 Independent Variables (without Mean Income)	23.87775083162673%	0.15426473655188488

Table 3: Mean Percentage Error (MPE) and Adjusted R^2 Values

→ Evaluation:

When we remove Population Size as an independent variable, there is lower fall in adjusted R^2 from 0.28 to 0.22 and lower increase in mean percentage error from 21.4 to 22.6%, compared to when we remove Mean Income where adjusted R^2 falls from 0.28 to 0.15 and MPE increased from 21.4 to 23.0. This shows how each independent variable has different correlation with dependent variable and how much it influences the dependent variable can be seen by the changes in adjusted R^2 value when we add or remove the dependent variable.

4 DISCUSSION

4.1 Alternative methods to find solution for Least squares problem

We can obtain the same solution for Least Squares Problem by using the following two methods.

Approach 1: (Lecture notes, page 109-111) Geometric interpretation

As mentioned before, we cannot find x so that $Ax = b$ is consistent, but we attempt to find $x = \hat{x}$ so that vector $A\hat{x}$ is "closest" to vector b (which means that $\|A\hat{x} - b\|$ is the minimum value of $\|Ax - b\|$.)

As we know, vector Ax is in the column space $V = \mathrm{Col}(A)$. Geometrically, $A\hat{x}$ is closest to b if and only if $A\hat{x}$ is the projection of b in subspace V containing $A\hat{x}$ (as shown in the figure below).

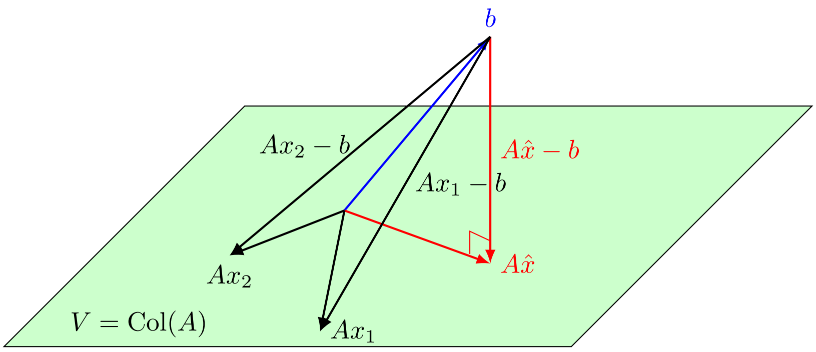

Figure 12: Geometric interpretation for Least Squares solution

Hence, $(A\hat{x} - b) \perp \mathrm{Col}(A)$, which is equivalent to $\langle a_j, (A\hat{x} - b)\rangle = 0$ or $a_j^T(A\hat{x} - b) = 0$ for all $j = 1, 2, \ldots, n$, where $A = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \end{pmatrix}$.

$$\rightarrow \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix} (A\hat{x} - b) = 0 \Leftrightarrow \boxed{A^T(A\hat{x} - b) = 0} \tag{4.1}$$

Approach 2: Derivation from calculus [6]

We have

$$f(x) = \|Ax - b\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}x_j - b_i \right)^2 \tag{4.2}$$

Take partial derivative with respect to x_k

$$\frac{\partial f(x)}{\partial x_k} = 2 \sum_{i=1}^{m} A_{ik} \left(\sum_{j=1}^{n} A_{ij}x_j - b_i \right) \tag{4.3}$$

We can notice that the above equation could be expressed in another way

$$\frac{\partial f(x)}{\partial x_k} = 2\begin{pmatrix} A_{1k} & A_{2k} & \dots & A_{mk} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} A_{1j}x_j - b_1 \\ \sum_{j=1}^{n} A_{2j}x_j - b_2 \\ \vdots \\ \sum_{j=1}^{n} A_{mj}x_j - b_m \end{pmatrix} \tag{4.4}$$

where $\begin{pmatrix} A_{1k} & A_{2k} & \dots & A_{mk} \end{pmatrix}$ is row k of the transpose matrix of A, A^T and

$$\begin{pmatrix} \sum_{j=1}^{n} A_{1j}x_j - b_1 \\ \sum_{j=1}^{n} A_{2j}x_j - b_2 \\ \vdots \\ \sum_{j=1}^{n} A_{mj}x_j - b_m \end{pmatrix} = Ax - b \tag{4.5}$$

Moreover, gradient of f is

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \dots \\ \frac{\partial f(x)}{\partial x_m} \end{pmatrix} = 2\begin{pmatrix} A_{11} & A_{21} & \dots & A_{m1} \\ A_{12} & A_{22} & \dots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{mn} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} A_{1j}x_j - b_1 \\ \sum_{j=1}^{n} A_{2j}x_j - b_2 \\ \vdots \\ \sum_{j=1}^{n} A_{mj}x_j - b_m \end{pmatrix} \tag{4.6}$$

where $\begin{pmatrix} A_{11} & A_{21} & \dots & A_{m1} \\ A_{12} & A_{22} & \dots & A_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{mn} \end{pmatrix} = A^T$

We can simplify (4.6) as

$$\nabla f(x) = 2A^T(Ax - b) \tag{4.7}$$

When $x = \hat{x}$, $f(x)$ is minimum and $\nabla f(\hat{x}) = 0$, thus

$$A^T(Ax - b) = 0 \tag{4.8}$$

$$\Leftrightarrow A^T Ax = A^T b \tag{4.9}$$

Furthermore, we can take inverse of $A^T A$ and obtain

$$\boxed{\hat{x} = (A^T A)^{-1} A^T b = A^+ b} \tag{4.10}$$

where $A^+ = (A^T A)^{-1} A^T$ is called the *pseudo-inverse* of a left-invertible matrix A

Comment: We obtain the same equation to find least squares solution as in approach 1.

$\rightarrow$ We can notice that both solutions (4.10) and (1.27) are the same, indeed, by substituting $A = QR = Q_1R_1$ into (4.10), we obtain

$$\begin{aligned}
\hat{x} &= (A^TA)^{-1}A^Tb && (4.11)\\
&= ((QR)^TQR)^{-1}(Q_1R_1)^Tb && (4.12)\\
&= (R^TQ^TQR)^{-1}R_1^TQ_1^Tb && (4.13)\\
&= (R^TR)^{-1}R_1^TQ_1^Tb && (4.14)\\
&= (R_1^TR_1)^{-1}R_1^TQ_1^Tb && (4.15)\\
&= R_1^{-1}(R_1^T)^{-1}R_1^TQ_1^Tb && (4.16)\\
&= R_1^{-1}Q_1^Tb && (4.17)
\end{aligned}$$

Note: $Q^TQ = I$; $R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix} \rightarrow R^T = \begin{pmatrix} R_1^T & 0^T \end{pmatrix} \rightarrow R^TR = R_1^TR_1$

4.2 Advantages and disadvantages of different methods:

(a) Gram-schmidt process:

- Inherently numerically unstable.
- While the application of the projections has an appealing geometric analogy to orthogonalization, the orthogonalization itself is prone to numerical error.
- A significant advantage however is the ease of implementation. In case, software does not have pre-built linear algebra library, this method seems to be useful for prototyping. However, software like matlab or python are already using other methods to do QR decomposition.

(b) Givens Rotation:

- The QR decomposition via Givens rotations is the most involved to implement, as the ordering of the rows required to fully exploit the algorithm is not trivial to determine.
- However, it has a significant advantage in that each new zero element a_{ij} affects only the row with the element to be zeroed (i) and a row above (i-1). This makes the Givens rotation algorithm more bandwidth efficient and parallelizable than the Householder reflection technique.

(c) Householder Reflection:

- The most simple of the numerically stable QR decomposition algorithms due to the use of reflections as the mechanism for producing zeroes in the R matrix.
- However, the Householder reflection algorithm is bandwidth heavy and not parallelizable, as every reflection that produces a new zero element changes the entirety of both Q and R matrices.

→ Rationale for using QR decomposition ($Rx = Q^T b$) instead of using normal equation ($A^T Ax = A^T b$) to solve least squares problem:

- By forming the product $A^T A$, we square the condition number of the problem matrix
- It is proved that using QR decomposition makes a better least squares estimate than the Normal equation in terms of solution quality (more numerically stable)

5 APPENDIX

Team members contribution

No.	Member	Area of contribution	Sign
1	Dinh Quang Huy	Theory	
2	Nguyen Duc Thang	Theory	Thắng
3	Trinh Duy Hieu	Discussion	
4	Albert Ariel Widiaatmaja	Implementation	Albert
5	Ammar B Anif	Example	
6	Ang Lin Min, Nicole	Example	

References

[1] "Solving least-squares with qr," https://johnwlambert.github.io/least-squares/#:~:text=By%20forming%20the%20product%20A,in%20terms%20of%20solution%20quality, accessed: 8/3/2022.

[2] "Qr matrix factorization," https://towardsdatascience.com/qr-matrix-factorization-15bae43a6b2, accessed: 8/3/2022.

[3] Wikipedia contributors, "Qr decomposition — Wikipedia, the free encyclopedia," https://en.wikipedia.org/w/index.php?title=QR_decomposition&oldid=1073018451, 2022, [Online; accessed 8-March-2022].

[4] c34klh123. Explore happiness data (2017, april 14). (accessed: April 14, 2022). [Online]. Available: https://www.kaggle.com/code/c34klh123/explore-happiness-data/data?select=2016.csv

[5] Gantalaswetha. Usa housing dataset linear regression (2019, may 14). (accessed: April 14, 2022). [Online]. Available: https://www.kaggle.com/code/gantalaswetha/usa-housing-dataset-linear-regression/data

[6] "Least squares approximation," http://www.seas.ucla.edu/~vandenbe/133A/lectures/ls.pdf, accessed: 12/3/2022.