Note: All vector spaces will be finite-dimensional vector spaces over the field $\mathbb{R}$.

1   Projections and the Gram-Schmidt Process

We begin with a review on projections, orthogonality, and the Gram-Scmidt process for finding orthogonal bases.

Definition 1.1. Let V be a finite dimensional vector space with decomposition $V = U \oplus W$. The **projection of V onto U along W** is the linear map $P_{U,W} : V \to U$ that assigns to each $v \in V$ the unique element $u \in U$ such that $v - u \in W$.

Exercise 1. Let U be a subspace of a finite dimensional vector space V. Show that a surjective linear map $P : V \to U$ is a projection map if and only if P is idempotent (i.e. $P^2 = P$).

Definition 1.2. A **finite dimensional inner product space** is a pair $(V, \langle \cdot, \cdot \rangle)$ consisting of a finite dimensional vector space V and an inner product $\langle \cdot, \cdot \rangle$ on V.

Exercise 2. Show that every finite dimensional vector space has an inner product (Use that the category of finite dimensional vector spaces over $\mathbb{R}$ with linear maps is equivalent to the category of Euclidean spaces $\mathbb{R}^n$ with linear maps).

Definition 1.3. Let $(V, \langle \cdot, \cdot \rangle)$ be finite dimensional inner product space and let $U \subseteq V$ be a subspace of V. The **orthogonal complement** of U in V is the subspace $U^\perp \subseteq V$ defined by:

$$U^\perp := \{v \in V \mid \langle v, \cdot \rangle|_U \equiv 0\}$$

Exercise 3. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and let U and W be subspaces of V. Show that:

1. $(U + W)^\perp = U^\perp \cap W^\perp$

2. $(U \cap W)^\perp = U^\perp + W^\perp$

Exercise 4. Let $(V, \langle \cdot, \cdot \rangle)$ and $(W, \langle \cdot, \cdot \rangle)$ be finite dimensional inner product spaces and let $A : V \to W$ be a linear transformation. Show that:

1. $\operatorname{im}(A^T)^\perp = \ker(A)$

2. $\operatorname{im}(A)^\perp = \ker(A^T)$

Definition 1.4. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and let U be a subspace of V. The **orthogonal projection** of V onto U is the projection $P_U : V \to U$ of V onto U along the orthogonal complement $U^\perp$ in the sense of Definition 1.1.

Exercise 5. Let U be a subspace of a finite dimensional inner product space $(V, \langle \cdot, \cdot \rangle)$ and let $U^\perp$ be its orthogonal complement.

1. Find a formula for the orthogonal projection of V onto U with respect to an orthogonal basis of U.

2. Show that $V = U \oplus U^\perp$

Definition 1.5. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and let U be a subspace of V. Given a vector $v \in V$, the **best approximation to v by vectors in U** is a vector that attains the minimum of the function:

$$f : U \to \mathbb{R}, \qquad f(u) = ||v - u||^2$$

where $|| \cdot ||$ is the norm on the vector space V induced by the inner product $\langle \cdot, \cdot \rangle$.

Exercise 6. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and let U be a subspace of V. Show that the best approximation of a vector $v \in V$ by vectors in U is the orthogonal projection $P_U(v)$ of the vector v onto the subspace U.

Exercise 7. Let A be an $m \times n$ matrix with real entries and $b \in \mathbb{R}^m$ a vector. Show that $Ax = b$ has a solution if and only if $b \in \ker(A^T)^\perp$. (Hint: The equalities $\text{im}(A^T) = \ker(A)^\perp$ and $\text{im}(A) = \ker(A^T)^\perp$ may be useful).

Theorem 1.6. (Gram-Schmidt Process.) Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and let $\{x_1, ... x_n\}$ be a basis of V. Define a set $\{v_1, ..., v_n\}$ of vectors in V as follows:

1. Let $v_1 := x_1$.

2. For each integer k such that $1 < k \le n$, let $V_k := \text{span}\{v_1, ..., v_{k-1}\}$.

3. For each integer k such that $1 < k \le n$, define $v_k := x_k - P_{V_k}(x_k)$, where $P_{V_k} : V \to V_k$ is the orthogonal projection of V onto V_k.

The set $\{v_1, ..., v_k\}$ is an orthogonal basis of V.

Exercise 8. Prove that the Gram-Schmidt Process works.

Exercise 9. Consider the basis $\{(1, 1, 0), (0, 1, 0), (1, 1, 1)\}$ of $\mathbb{R}^3$. Use the Gram-Schmidt process to obtain an orthonormal basis of $\mathbb{R}^3$ starting from this one.

Exercise 10. Use the Gram-Schmidt Process to find an orthogonal basis of the image of the matrices:

1. $\begin{pmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{pmatrix}$

2. $\begin{pmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{pmatrix}$

Exercise 11. For any integer $k \ge 0$, let $\mathbb{R}[k]$ be the space of polynomials of degree at most k with real coefficients. Fix integers $1 \le k \le n$.

1. Fix distinct numbers $x_0, x_1, ..., x_n$, where k is an integer such that $1 \le k \le n$, and define the following bilinear function on $\mathbb{R}[k] \times \mathbb{R}[k]$:

$$\langle p, q \rangle := p(x_0)q(x_0) + p(x_1)q(x_1) + ... + p(x_n)q(x_n)$$

   Show that this is an inner product.

2. Let $\mathbb{R}[4]$ be space of polynomials of degree at most 4 with real coefficients. Consider the inner product on this space given by the formula in part (1) for the numbers $-2, -1, 0, 1, 2$. View the space of polynomials $\mathbb{R}[2]$ of degree at most 2 with real coefficients as a subspace of the space $\mathbb{R}[4]$. Use this inner product and the Gram-Schmidt process to find an orthogonal basis of the subspace $\mathbb{R}[2]$ starting from the basis $\{1, t, t^2\}$.

2   The QR Factorization

Exercise 12. Let A be an $m \times n$ matrix. Suppose $A = QR$ where Q and R are matrices of dimensions $m \times n$ and $n \times n$ respectively.

1. Suppose A has linearly independent columns. Show R is invertible. (Hint: Consider the subspace $\ker R$ of $\mathbb{R}^n$.)

2. Suppose R is invertible. Show that $\mathrm{im}(A) = \mathrm{im}(Q)$.

Definition 2.1. Given an $m \times n$ matrix A, a QR **factorization** of the matrix A is a factorization of the form $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis of $\mathrm{im}(A)$ and R is an $n \times n$ upper triangular matrix with positive entries on its diagonal.

Theorem 2.2. If A is an $m \times n$ matrix with linearly independent columns, then A has a QR factorization.

Exercise 13. (QR **Factorization via the Gram-Schmidt Process.**) Let A be an $m \times n$ matrix A that has linearly independent columns. Prove Theorem 2.2 by constructing a QR factorization of A using the Gram-Schmidt Process.

Exercise 14. Compute the QR factorization of the following matrices using the Gram-Schmidt Process:

1. $\begin{pmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{pmatrix}$

2. $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$

Definition 2.3. Let A be an $m \times n$ matrix with linearly independent columns. If $m \ge n$, the **full** QR **factorization** of the matrix A is a factorization of the form:

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$$

where Q is an $m \times m$ orthogonal matrix and R is an $n \times n$ invertible matrix. If $m \leq n$, the **full QR factorization** takes the form:

$$A = Q \begin{pmatrix} R & 0 \end{pmatrix}$$

Exercise 15. Given an $m \times n$ matrix A, show how to obtain a QR factorization from a full QR factorization. Show how to obtain a full QR factorization from a QR factorization.

Definition 2.4. A **Householder matrix** or **elementary reflector** is a matrix of the form:

$$H = \mathrm{Id} - 2\mathbf{u}\mathbf{u}^{\mathbf{T}}$$

where $\mathbf{u}$ is a unit vector (viewed as a column vector).

Exercise 16. Let H be a Householder matrix. Prove the following properties of Householder matrices:

1. Prove that $H^2 = \mathrm{Id}$.

2. Prove that H is an orthogonal matrix.

3. Prove that if H is an $n \times n$ matrix, then H has eigenvalues 1 and -1 with multiplicities $n - 1$ and 1 respectively. (Hint: What does $H = \mathrm{Id} - \mathbf{u}\mathbf{u}^{\mathbf{T}}$ do to vectors orthogonal to the vector $\mathbf{u}$? How many such vectors are there?)

4. Prove the determinant of H is -1.

5. Prove that $H = \mathrm{Id} - \mathbf{u}\mathbf{u}^{\mathbf{T}}$ corresponds to reflection about the hyperplane in $\mathbb{R}^n$ defined by the unit vector $\mathbf{u}$. One way to check this is to show that the linear transformation $x \mapsto Hx$ satisfies that the orthogonal projection of x and Hx onto the hyperplane defined by $\mathbf{u}$ is the same.

6. Provide examples in 2 and 3 dimensions of Householder matrices that illustrate the reflection nicely.

Exercise 17. Let A be an $m \times n$ matrix and let $\mathbf{a}$ be the first column of A. Let $\mathbf{v}$ be the vector:

$$\mathbf{v} := \mathbf{a} - ||\mathbf{a}||\mathbf{e_1}$$

where $\mathbf{e_1}$ is the first standard basis vector of $\mathbb{R}^m$, and let $\mathbf{u}$ be the corresponding unit vector $\mathbf{u} := \mathbf{v}/||\mathbf{v}||$. Define the Householder matrix $H := \mathrm{Id} - 2\mathbf{u}\mathbf{u}^{\mathbf{T}}$. Then:

1. Compute $H\mathbf{a}$. What can you say about the product HA?

2. Interpret H geometrically in terms of the subspaces $\mathrm{span}(\mathbf{a})$ and $\mathrm{span}(\mathbf{e_1})$

Exercise 18. (QR **Factorization via Hausholder transformations.**) Let A be an $m \times n$ matrix with linearly independent columns. Consider the following algorithm:

1. For step 1, let $\mathbf{a_1}$ be the first column of A and form the Householder matrix described in the previous exercise with $\mathbf{a} = \mathbf{a_1}$. Call it H_1' and set $H_1 := H_1'$. Form the product $H_1 A$.

2. For the kth step, let A_k be the matrix obtained by deleting the first k rows and columns of the matrix:

$$H_{k-1}H_{k-2}\cdots H_1 A$$

Form the Householder matrix as in step 1, but for the matrix A_k. Call it H'_k and define:

$$H_k := \begin{pmatrix} \text{Id}_{k-1} & 0 \\ 0 & H'_k \end{pmatrix}$$

Now form the product $H_2 H_1 A$.

3. Repeat the process for $N := \min\{m - 1, n\}$ steps and then form the matrix:

$$\widetilde{R} := H_N H_{N-1} \cdots H_1 A$$

Now do the following:

1. Show how to obtain a QR decomposition starting from the matrix $\widetilde{R}$. (Hint: This matrix gives the right hand side matrix of the full QR decomposition).

2. Geometrically describe the difference between the algorithms producing the QR decomposition based on the Gram-Schmidt process and the Householder reflections.

Exercise 19. Compute the QR factorization of the following matrix using Hausholder transformations:

$$\begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}$$

Exercise 20. Compute the QR factorization of the matrices in exercise 14.

Exercise 21. Let A be an $m \times n$ matrix with linearly independent columns and let $A = QR$ be a corresponding QR factorization. Let p be an integer such that $1 \le p \le n$. Partition A into the form:

$$A = (A_1 \; A_2)$$

where A_1 is an $m \times p$ matrix and A_2 is an $m \times (n - p)$ matrix. Give an algorithm for obtaining a QR factorization of A_1.

3 Least-Squares

It's very common for linear systems that show up in applications to **not** have a solution. One can still ask for the closest "approximate solution".

Definition 3.1. Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$ an m-dimensional vector. A **least-squares solution** of the system $Ax = b$ is a vector x_m attaining the minimum of the function:

$$f : \mathbb{R}^n \to \mathbb{R}, \qquad f(x) := ||b - Ax||^2$$

The quantity $f(x_m) = ||b - Ax_m||$ is known as the **least-squares error** of the least-squares solution x_m. We'll denote the set of least-squares solutions by $\text{LS}(A, b)$.

Definition 3.2. Let A be an $m \times n$ matrix and let $b \in \mathbb{R}^m$, the **normal equation** associated to the system $Ax = b$ is the system $A^T Ax = A^T b$.

Exercise 22. Let A be an $m \times n$ matrix and let $b \in \mathbb{R}^m$. Let $\text{LS}(A, b)$ be the set of least-squares solutions. Show that:
$$\text{LS}(A, b) = \left\{ x \in \mathbb{R}^n \mid A^T Ax = A^T b \right\}$$
Explain geometrically the relationship between least-squares solutions and orthogonal projections.

Exercise 23. Compute a least-squares solution of the following $Ax = b$ system:
$$A = \begin{pmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{pmatrix} \qquad b = \begin{pmatrix} 3 \\ 1 \\ -4 \\ 2 \end{pmatrix}$$

Exercise 24. Describe the least squares solutions of the following $Ax = b$ systems:

1. $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 3 \\ 8 \\ 2 \end{pmatrix}$

2. $A = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 5 \\ 1 \\ 0 \end{pmatrix}$

Exercise 25. Let A be an $m \times n$ matrix. Show that:

1. $\ker A = \ker A^T A$. (Hint: For the inclusion $\supseteq$ consider the vector $x^T A^T Ax$ for arbitrary $x \in \ker(A^T A)$.)

2. Show $\text{rank} A^T A = \text{rank} A$.

3. If $A^T A$ is invertible, then A has linearly independent columns.

Theorem 3.3. (**Conditions for Uniqueness of Least-Squares Solutions.**) Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. The following are all equivalent:

1. There is only one least-squares solution, i.e. the set of least-squares solutions $\text{LS}(A)$ is a singleton.

2. The columns of A are invertible.

3. The matrix $A^T A$ is invertible.

Exercise 26. Prove Theorem 3.3. (Hint: For one of the implications it may be useful to show that $\ker A = \ker A^T A$.)

Exercise 27. Let A be an $m \times n$ matrix with linearly independent columns and let $b \in \mathbb{R}^m$ be a vector. Use a QR factorization of A to find the least-squares solutions to the system $Ax = b$.

Exercise 28. Suppose there is a unique solution to a least-squares problem $Ax = b$. Let $c \in \mathbb{R}$ be a nonzero scalar. What is the set of least-squares solutions $\mathrm{LS}(A, cb)$ in terms of the least-squares solution of $Ax = b$?

Exercise 29. (**Rayleigh Quotients**) Let A be an $n \times n$ matrix. Suppose $v \in \mathbb{R}^n$ is an *approximate* eigenvalue. That is, suppose the system:

$$Av = \lambda v$$

does not have a solution $\lambda \in \mathbb{R}$, but "almost" has one. Rephrase the system so that you can use the least-squares method to best-approximate the eigenvalue λ. This estimate is called a **Rayleigh quotient**.

4 Linear Models: Regression

Definition 4.1. A **general linear model** for a relationship between $y \in \mathbb{R}$ and $x \in \mathbb{R}^m$ is an equation of the form:

$$y = \beta_0 + \beta_1 f_1(x) + \beta_2 f_2(x) + \ldots + \beta_k f_k(x) \tag{4.1}$$

Given a data set $\{(x_i, y_i)\}_{i=1,\ldots,n}$ consisting of observations of the variables $x \in \mathbb{R}^m$ and $y \in \mathbb{R}$, an approach to model the underlying phenomenon is to suppose the variables x and y satisfy a general linear model as in Definition 4.1. One then tries to find parameters $\widehat{\beta} = \left(\widehat{\beta}_0, \ldots, \widehat{\beta}_k \right)$ such that the variables satisfy the corresponding linear model up to some small error. That is one could try to solve the system:

$$\begin{pmatrix} 1 & f_1(x_1) & f_2(x_1) & \ldots & f_k(x_1) \\ 1 & f_1(x_2) & f_2(x_2) & \ldots & f_k(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & f_1(x_n) & f_2(x_n) & \ldots & f_k(x_n) \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{4.2}$$

The matrix in equation (4.2) is usually denoted by X and is called the **design matrix**, the vector $\beta = (\beta_0, \ldots, \beta_k)$ is called the **parameter vector**, and the vector $y = (y_1, \ldots, y_n)$ is called the vector of **observations**. Thus, the system (4.2) becomes $X\beta = y$.

Note that in most cases an exact solution to (4.2) doesn't exist. In such a case, one seeks the best approximation, and this is often done via least-squares optimization. The error in this context is often called the **residual vector** and denoted by ϵ.

Exercise 30. (**Lines of regression.**) The most basic example of (4.2) comes from **lines of regression** or **least-squares lines**. Suppose you are given data $\{(x_i, y_i)\}_{i=1,\ldots,n}$ where $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$. Model this data via the following linear model:

$$y = \beta_0 + \beta_1 x$$

1. Give the design matrix for this model.

2. Let $\widehat{\beta} = \left(\widehat{\beta}_0, \widehat{\beta}_1 \right)$ be the least-squares solutions of the resulting system. Geometrically interpret the solution and its relationship to the data points $\{(x_i, y_i)\}$.

3. Geometrically interpret the residual vector ϵ.

4. Do you think this method is appropriate for all data? Justify if yes. If not, provide an example of data for which it isn't.

Exercise 31. Find the least-squares line that best fits the data: $(2,3), (3,2), (5,1)$, and $(6,0)$.

Exercise 32. Given a data set $\{(x_i, y_i)\}_{i=1,...,n}$, show that there is a unique least-squares line if the data contain at least two data points with distinct x values.

Exercise 33. Given a data set $\{(x_i, y_i)\}_{i=1,...,n}$, let $\bar{x}$ be the mean of the x-values. Define the variable $x^* := x - \bar{x}$ and consider the data set $\{(x_i^*, y_i)\}_{i=1,...,n}$. This data set is said to be in **mean-deviation form**.

1. Show that the corresponding design matrix is an orthogonal matrix.

2. How does this affect solving the normal equation for the corresponding least-squares problem?

Exercise 34. Given a data set $\{(x_i, y_i)\}_{i=1,...,n}$. The data points are not always equally reliable (e.g. they might be drawn from populations with probability distributions with different variances). It might thus be prudent to give each data point a different weight. More precisely, one can define the inner product on $\mathbb{R}^n$:

$$\langle u, v \rangle := w_1^2 u_1 v_1 + w_2^2 u_2 v_2 + ... + w_n^2 u_n v_n$$

and perform the linear regression with respect to this inner product.

1. Give the normal equation for the least-squares problem with respect to this inner-product (Hint: Form a matrix with the weights).

2. Give the corresponding least-square error. This is often called the **weighted sum of squares for error**.

Exercise 35. Given a data set $\{(x_i, y_i)\}_{i=1,...,n}$ it is very common for the data to *not* have a linear trend. In such a case a linear relationship between the x and y may not be the best guess for the functional relationship $y = f(x)$. One could instead guess, for example, a quadratic relationship:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Note: We are still working with the general linear model from Definition 4.1. The "linearity" of the general linear model is linearity in the parameters β_j not the relationship between the variables x and y.

1. Give the design matrix for the above quadratic relationship $y = f(x)$.

2. Give the design matrix for a cubic relationship $y = f(x)$.

3. Give the design matrix for a general polynomial relationship $y = f(x)$ of degree k.

Exercise 36. Consider the data $(1, 1.8), (2, 2.7), (3, 3.4), (4, 3.8), (5, 3.9)$.

1. Plot the data.

2. Use an appropriate general linear model to fit the data.

Exercise 37. (This is a really nice short problem found in Lay's book [L94] based on Gauss' original work on regression and the prediction of the orbit of the asteroid Ceres.) Ignoring the gravitational attraction of the planets, Kepler's first law predicts a comet should have an elliptic, parabolic, or hyperbolic orbit. In suitable polar coordinates (r, θ) the comet should obey:

$$r = \beta + e(r\cos\theta)$$

where β is some constant and e is the eccentricity of the orbit. Suppose you have the observations $(3, .88), (2.3, 1.1), (1.65, 1.42), (1.25, 1.77), (1.01, 2.14)$.

1. Use a least-squares approach to determine the parameters β and e.

2. Determine the type of orbit from the parameter e. If $0 \le e < 1$, the orbit is an ellipse, if $e = 1$ it is a parabola, and if $e > 1$ it is a hyperbola.

3. Predict the position r when $\theta = 4.6$.

Exercise 38. Let $\{(t_i, y_i)\}_{i=1,...,n}$ be a set of data where t is time and the data exhibits seasonal fluctuations. Posit a model for this data and give its design matrix.

Exercise 39. Suppose you are given a data set $\{(u_i, v_i, y_i)\}_{i=1,...,n}$ and you posit there is a relationship $y = f(u, v)$.

1. Suppose the relationship is given by a plane. Give a general linear model for this dependence and the corresponding design matrix. This generalizes lines of regression to **planes of regression**.

2. Suppose there is periodic dependence on the u variable, quadratic dependence on the v variable, and linear dependence on the product uv. Give a general linear model for this relationship and the corresponding design matrix.

Exercise 40. (Interpolating polynomials.) Let $x_1, ..., x_n$ be numbers. A **Vandermonde matrix** is the design matrix for the general linear model given by:

$$f_j(x) = x^j \qquad j = 0, ..., n-1$$

1. Write down the general form of the Vandermonde matrix.

2. Let V be the Vandermonde matrix for the numbers $x_1, ..., x_n$ and let $y \in \text{im}(V)$. Let $c = (c_0, ..., c_{n-1})$ be such that $Vc = y$ and define the polynomial:

$$p(x) := c_0 + c_1 x + c_2 x^2 + ... + c_{n-1} x^{n-1}$$

Show that the graph of p contains the points $\{(x_i, y_i)\}_{i=1,...,n}$. That is, the polynomial p interpolates the data.

3. Show that if the numbers x_i, $i = 1, ..., n$, are distinct then the rank of the matrix V is n. (Hint: Show the columns are linearly independent by considering the roots of the polynomial p.)

4. Show that if the data set $\{(x_i, y_i)\}_{i=1,\dots,n}$ is such that the x_i are distinct then one can always find an interpolating polynomial of degree $n-1$.

Exercise 41. Find a polynomial interpolating $(-2, 3), (-1, 5), (0, 5), (1, 4),$ and $(2, 3)$.

Exercise 42. (Trend analysis.) It is not always clear what is the trend of some given data $\{(x_i, y_i) \in \mathbb{R}^2\}_{i=1,\dots,n}$. One way is to use the data points to perform **trend analysis**.

1. Compute an orthogonal basis $\{p_0, p_1, p_2\}$ of $\mathbb{R}[2]$ with respect to the inner product:

$$\langle p, q \rangle := p(x_0)q(x_0) + p(x_1)q(x_1) + \dots + p(x_n)q(x_n)$$

by performing the Gram-Schmidt Process starting from the basis $\{1, t, t^2\}$.

2. Let $\{p_0, p_1, p_2\}$ be an orthogonal basis of $\mathbb{R}[2]$ with respect to the inner product:

$$\langle p, q \rangle := p(x_0)q(x_0) + p(x_1)q(x_1) + \dots + p(x_n)q(x_n)$$

Let f be a polynomial in $\mathbb{R}[n]$ such that $f(x_i) = y_i$ for $i = 1, \dots, n$ (see Exercise 40). Let $P : \mathbb{R}[n] \to \mathbb{R}[2]$ be the orthogonal projection with respect to the inner product on $\mathbb{R}[n]$. The projection of the interpolating polynomial f gives a **quadratic trend function** fitted to the data:

$$P(f) = c_0 p_0 + c_1 p_1 + c_2 p_2$$

Show that knowledge of the polynomial f is not needed to obtain the projection $P(f)$, it suffices to know the values y_i.

3. Consider the data points $(-2, 3), (-1, 5), (0, 5), (1, 4),$ and $(2, 3)$.

   (a) Fit a linear trend function to the data.

   (b) Fit a quadratic trend function to the data.

References

[Hal58]   P.R. Halmos. Finite-Dimensional Vector Sapces. Reprinting of the 1958 second edition. *Undergraduate Texts in Mathematics.* Springer-Verlag, New York-Heidelberg, 1974.

[L94]   D.C. Lay. Linear Algebra and its Applications. Fourth Edition. Addison-Wesley, 2012.

[Rom08]   S. Roman. Advanced Linear Algebra. Third Edition. *Graduate Texts in Mathematics, 135.* Springer, New York, 2008.