Linear Algebra Working Group :: Day 1

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 of 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 $P_{U,W} : V \to V$ that assigns to each $v \in V$ the unique element $u \in U$ such that $v - u \in W$.

Exercise 1. Let V be a finite dimensional vector space with decomposition $V = U \oplus W$, and suppose that $P_{U,W} : V \to V$ is the projection of V onto U along W. Show that $P_{U,W}$ has image precisely U and kernel W.

Exercise 2. Let V be a finite dimensional vector space. Show that a linear map $P : V \to V$ is a projection map onto its image if and only if P is idempotent (i.e. $P^2 = P$). (*Hint: Use the rank-nullity theorem.*)

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 3. Show that every finite dimensional vector space has an inner product (*Hint: Use a basis.*)

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 4. Let $(V, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space. Show that for all linear subspaces U of V we have that $(U^\perp)^\perp = U$.

Exercise 5. 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. $(U \cap W)^\perp = U^\perp + W^\perp$

Definition 1.4. Let $(V, \langle \cdot, \cdot \rangle_V)$ and $(W, \langle \cdot, \cdot \rangle_W)$ be finite dimensional inner product spaces and let $A : V \to W$ be a linear transformation. The **transpose** or **adjoint** of the map A with respect to the inner products is the unique linear transformation $A^T : W \to V$ such that:

$$\langle A(v), w \rangle_W = \left\langle v, A^T(w) \right\rangle_V, \qquad \text{for all } v \in V, w \in W.$$

Exercise 6. Let $(V, \langle \cdot, \cdot \rangle_V)$ and $(W, \langle \cdot, \cdot \rangle_W)$ be finite dimensional inner product spaces and let $A : V \to W$ be a linear transformation. Show that $(A^T)^T = A$.

Exercise 7. 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. $\operatorname{im}(A)^\perp = \ker(A^T)$

Exercise 8. 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 $b \in \ker(A^T)^\perp$.

Definition 1.5. 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 $P_U : V \to V$ of V onto U along the orthogonal complement $U^\perp$ in the sense of Definition 1.1.

Exercise 9. Let $(V, \langle\cdot,\cdot\rangle)$ be finite dimensional inner product space, let U be a subspace, and let $P_U : V \to V$ be the orthogonal projection onto U. Sketch and prove the Pythagorean identity:

$$||v - u||^2 = ||v - P_U(v)||^2 + ||P_U(v) - u||^2, \qquad \text{for any } v \in V, u \in U,$$

where $||\cdot||$ is the norm induced by the inner product.

Exercise 10. 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. Let $P_U : V \to V$ be the orthogonal projection onto U, and suppose that $u_1, \dots, u_k$ is an orthogonal basis of U. Show that for all vectors $v \in V$:

$$P_U(v) := \sum_{i=1}^{k} \frac{\langle v, u_i\rangle}{\langle u_i, u_i\rangle} u_i.$$

Definition 1.6. 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 11. Let $(V, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner product space and let U be a subspace of V. Let k be the dimension of U and n be the dimension of V. Show that the best approximation of a vector $v \in V$ by vectors in U is the orthogonal projection $P_U(v)$. *Hint: Use exercise 9, or use the formula of exercise 10 and calculus.*

Exercise 12. Let $(V, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner product space. Suppose that u and v are orthogonal vectors. Show that u and v must be linearly independent.

Theorem 1.7. (**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 $1 < k \leq n$, let $V_k := \operatorname{span}\{v_1, ..., v_{k-1}\}$.
3. For each integer k such that $1 < k \leq n$, define $v_k := x_k - P_{V_k}(x_k)$, where $P_{V_k} : V \to V$ is the orthogonal projection of V onto V_k.

4. Continue until you obtain n vectors $\{v_1, \ldots, v_n\}$.

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

Exercise 13. Sketch an explanation of how the Gram-Schmidt process works.

Exercise 14. Prove that the Gram-Schmidt Process works.

Exercise 15. Let $\{x_1, \ldots, x_n\}$ be a basis of $\mathbb{R}^n$, and let $\{v_1, \ldots, v_n\}$ be the basis obtained via the Gram-Schmidt process.

1. Show that for all $1 \le k \le n$, we have that:

$$\operatorname{span}\{x_1, \ldots, x_k\} = \operatorname{span}\{v_1, \ldots, v_k\}.$$

2. Use this to say that for each $1 \le i < n$, some of the weights in:

$$x_i = r_{1,i}v_1 + r_{2,i}v_2 + \cdots + r_{n,i}v_n \tag{1.1}$$

   must be zero.

3. What can you say about the weight r_{ii} in (1.1)?

Exercise 16. 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 17. Use the Gram-Schmidt Process to find an orthogonal basis of the image of the matrix:

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

Exercise 18. For any fixed integer $k \ge 0$, let $\mathbb{R}[k]$ be the space of polynomials of degree at most k with real coefficients. This is a finite dimensional vector space under pointwise addition and scalar multiplication.

1. Give a basis for the space of polynomials $\mathbb{R}[k]$.

2. Fix distinct numbers $x_1, x_2, ..., x_{n+1}$, and define the following bilinear function on the vector space $\mathbb{R}[n] \times \mathbb{R}[n]$:

$$\langle p, q\rangle := p(x_1)q(x_1) + p(x_2)q(x_2) + ... + p(x_{n+1})q(x_{n+1})$$

   Show that this is an inner product. (*Hint: To check definiteness, note that the equality* $\langle p, p\rangle = 0$ *says something about the roots of the polynomial* p.)

3. 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 19. 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 20. (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 21. Compute the QR factorization of the following matrix using the Gram-Schmidt Process:

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

Definition 2.3. Let A be an $m \times n$ matrix. If $m \geq n$, 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 22. 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 23. 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 $\mathbf{u}$. Since H is orthogonal, 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 24. OPTIONAL: (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 k^{th} 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:
$$H_k := \begin{pmatrix} \mathrm{Id}_{k-1} & 0 \\ 0 & H_k' \end{pmatrix}$$
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 25. Compute the QR factorization of the following matrix using Hausholder transformations:
$$\begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}$$

3 Least-Squares

Remark 3.1. It's very common for linear systems that show up in applications to **not** have a solution. One can still ask for the "best approximate solution". This happens, for example, when we have an $m \times n$ matrix A and we want to solve a system:

$$Ax = b, \qquad x \in \mathbb{R}^n,$$

where $m > n$. For example, the rows of A could be m measurements of n parameters in some experimental context. Such a system does not have a solution for general $b \in \mathbb{R}^m$.

Definition 3.2. Let V and W be finite-dimensional inner product spaces, let $A : V \to W$ be a linear map, and let $b \in W$ be a vector. A **least-squares solution** of the equation $A(x) = b$ is a vector $x_m \in V$ attaining the minimum of the function:

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

where $|| \cdot ||$ is the norm induced on W by the inner product. The quantity $\varepsilon(x_m) = ||b - Ax_m||^2$ is known as the **least-squares error** of the least-squares solution x_m. We'll denote the subset of V consisting of least-squares solutions by $\mathrm{LS}(A, b)$.

Exercise 26. Relate the set of least squares solution of Definition 3.2 with an orthogonal projection in the sense of Definition 1.5. That is, define a set $\mathrm{PR}(A, b)$ of solutions to an equation involving an orthogonal projection such that $\mathrm{LS}(A, b) = \mathrm{PR}(A, b)$.

Definition 3.3. Let $A : V \to W$ be a linear map of inner product spaces, let $b \in W$ be a vector, and let $A^T : W \to V$ be the adjoint map as in Definition 1.4. The **normal equation** associated to the equation $A(x) = b$ is the equation $A^T A(x) = A^T b$. We will denote the subset of V consisting of solutions to the normal equation by $\mathrm{NS}(A, b)$.

Exercise 27. Let $A : V \to W$ be a linear map between inner product spaces, and let $b \in W$ be a vector. Let $\mathrm{LS}(A, b)$ be the set of least-squares solutions of the equation $A(x) = b$, and let $\mathrm{NS}(A, b)$ be the set of solutions of the normal equation of Definition 3.3. Show that:

$$\mathrm{LS}(A, b) = \mathrm{NS}(A, b).$$

Exercise 28. Compute a least-squares solution of the following $Ax = b$ system:

$$A = \begin{pmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{pmatrix} \qquad b = \begin{pmatrix} 3 \\ 1 \\ -4 \\ 2 \end{pmatrix}$$

(*Hint: Use the normal equation.*)

Exercise 29. 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 30. Let $A : V \to W$ be a linear map between inner product spaces and let $A^T : W \to V$ be its adjoint. Show that:

1. $\ker A = \ker A^T A$. (*Hint: For the inclusion* $\supseteq$ *use the definition of the adjoint.*)

2. Show $\dim \operatorname{im} A^T A = \dim \operatorname{im} A$.

Theorem 3.4. (**Conditions for Uniqueness of Least-Squares Solutions.**) Let $A : V \to W$ be a linear map between inner product spaces. The following are all equivalent:

1. There is only one least-squares solution in $\mathrm{LS}(A, b)$ for each fixed vector $b \in W$.

2. The kernel of A is trivial.

3. The endomorphism $A^T A : V \to V$ is invertible.

Remark 3.5. Theorem 3.4 is often stated using matrices. In such a case, the second statement is often given as the equivalent assertion that the columns of A are linearly independent. This is one way of checking it in practice.

Exercise 31. Prove Theorem 3.4. (*Hint: Use exercises 26 and 30*)

Exercise 32. Let A be an $m \times n$ matrix with linearly independent columns, let $A = QR$ be its QR factorization, and let $b \in \mathbb{R}^m$ be a vector.

1. Show that the solution to the equation $Rx = Q^T b$ is the same as the least squares solution.

2. Why is the equation $Rx = Q^T b$ desirable to find a solution? Can you write the solution to this equation explicitly? Given the QR factorization, how "easy" is it to solve this equation? (*Hint: It's all about the particular form of* R.)

Definition 3.6. Let $A : V \to V$ be a self-adjoint (i.e. $A^T = A$) linear endomorphism of an inner product space V, and let $v \in V$ be a nonzero vector. The **Rayleigh quotient** of A and v is the quantity:

$$r(v) := \frac{\langle v, Av \rangle}{\langle v, v \rangle},$$

where $\langle \cdot, \cdot \rangle$ is the inner product on V.

Exercise 33. (**Rayleigh Quotients**) Let $A : V \to V$ be a self-adjoint linear endomorphism of an inner product space, and let $v \in V$ be a nonzero vector. Obtain the Rayleigh quotient as a best approximation to the problem of finding an *approximate* eigenvalue of A corresponding to the "eigenvector" v. That is, find the Rayleigh quotient as the best approximation to the equation:

$$Av = \lambda v$$

in the unknown real variable λ.

Remark 3.7. Rayleigh quotients are used in the eigenvalue numerical estimation method called *Rayleigh quotient iteration*. This is an iterative method, which yields succesive guesses of an eigenvalue starting from an initial guess of eigenvalue and eigenvector.

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:

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

where $f_i : \mathbb{R}^m \to \mathbb{R}$ are functions and $\beta = (\beta_1, \ldots, \beta_k) \in \mathbb{R}^k$ are the **parameter variables**.

A general linear model as in Definition 4.1 is linear in the sense that it is linear in the parameters β. 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. That is, one could try to solve the system:

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

to find the parameters β.

Definition 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 many cases there are many more observations than parameters (i.e. n is much larger than k), so an exact solution to (4.2) doesn't exist in general. In such a case, one seeks the best approximation, and this is often done via least-squares optimization. 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.

Definition 4.3. The least squares error associated to least-squares solution of the system (4.2) is called the **residual vector** and denoted by ϵ.

Exercise 34. (**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 $(x_i, y_i) \in \mathbb{R}^2$. Set $x := (x_1, \ldots, x_n)$ and $y := (y_1, \ldots, y_n)$. 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 a rough example of data points for which it isn't.

Exercise 35. Find the least-squares line that best fits the data: $(2, 3), (3, 2), (5, 1)$, and $(6, 0)$.

Exercise 36. Given a data set $\{(x_i, y_i)\}_{i=1,\dots,n}$, show that there is a unique least-squares line if the data contain at least two data points with distinct x values.

Exercise 37. Given a data set $\{(x_i, y_i)\}_{i=1,\dots,n}$, let $\overline{x}$ be the mean of the x-values. Define the variable $x^* := x - \overline{x}$ and consider the data set $\{(x_i^*, y_i)\}_{i=1,\dots,n}$. This data set is said to be in **mean-deviation form**.

1. Show that the corresponding design matrix for least squares using $\{(x_i^*, y_i)\}_{i=1,\dots,n}$ is such that $X^T X = I$ and $XX^T = I$, i.e. the columns are orthogonal.
2. How does this affect solving the normal equation for the corresponding least-squares problem?

Exercise 38. Given a data set $\{(x_i, y_i)\}_{i=1,\dots,n}$. 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 $\mathbb{R}^n$:

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

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 39. Given a data set $\{(x_i, y_i)\}_{i=1,\dots,n}$ 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 40. 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 41. (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 $0 \leq e < 1$, 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 42. 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 43. 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 44. (**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 $\{(x_i, y_i)\}_{i=1,...,n}$ is such that the x_i are distinct then one can always find an interpolating polynomial of degree $n - 1$.

Exercise 45. Find a polynomial interpolating $(-2,3),(-1,5),(0,5),(1,4)$, and $(2,3)$.

Exercise 46. (**Trend analysis.**) It is not always clear what is the trend of some given data $\{(x_i, y_i) \in \mathbb{R}^2\}_{i=1,\ldots,n}$. 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:

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

   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:

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

   Let f be a polynomial in $\mathbb{R}[n]$ such that $f(x_i) = y_i$ for $i = 1, ..., n$ (see Exercise 44). 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:

$$P(f) = c_0p_0 + c_1p_1 + c_2p_2$$

   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.