1 Recap

1.1 Orthogonal Decomposition Revisited

Given a matrix $A \in \mathbb{R}^{m \times n}$. Any vector $b \in \mathbb{R}^m$ can be uniquely expressed as $b = x + y$ for which $x \in C(A)$ and $y \in N(A^\top)$. In particular, $x$ and $y$ are the orthogonal projections of $b$ onto $C(A)$ and $N(A^\top)$ respectively.

When $A$ is tall ($m \geq n$) and has linearly independent columns, we can write

$$x = A(A^\top A)^{-1}A^\top b \quad \text{and} \quad y = (I - A(A^\top A)^{-1}A^\top)b.$$ 

In other words, the projection matrix onto $C(A)$ is $P = A(A^\top A)^{-1}A^\top$ and onto $N(A^\top) = Q = I - P$.

It is a good exercise to verify that such decomposition makes sense; that is to explain why $(A^\top A)^{-1}$ exists, $x \in C(A)$, $y \in N(A^\top)$, and $b = x + y$.

1.2 Inverses

Given a tall matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and linearly independent columns; i.e. $\text{rank}(A) = n$. Then $A$ has a left inverse $(A^\top A)^{-1}A^\top$.

Given a wide matrix $A \in \mathbb{R}^{m \times n}$ with $m \leq n$ and linearly independent rows; i.e. $\text{rank}(A) = m$. Then $A$ has a right inverse $A^\top(AA^\top)^{-1}$.

1.3 Overdetermined System

Consider the linear system $Ax = b$, where we have more equations than variables; i.e. $A$ is tall with more rows than columns. The system may not have a solution that satisfies all equations.

Least Squares Approximate Solution: Assume linearly independent columns

1. Orthogonal projection: Project $b$ onto the column space of $A$, i.e. $\tilde{b} = \text{proj}_{C(A)}b = A(A^\top A)^{-1}A^\top b$. Then solve $Ax = \tilde{b} \Rightarrow x = (A^\top A)^{-1}A^\top b$. This vector minimizes the norm of the residual $r = Ax - b$.

2. Left Inverse: Consider the (specific) left inverse $B = (A^\top A)^{-1}A^\top$. Then approximate $x = Bb = (A^\top A)^{-1}A^\top b$.

3. Optimization: Want to find $x$ that minimizes $\|Ax - b\|^2$. Setting gradient to zero gives $x = (A^\top A)^{-1}A^\top b$.

All these approaches give the same result – the least-squares approximate solution is $x_* = (A^\top A)^{-1}A^\top b$. 
1.4 Underdetermined System

More variables than equations; i.e. \( A \) has more columns than rows. The system has infinitely many solutions, and we need to pick a specific one.

**Minimum Norm Solution:** Assuming linearly independent rows, we pick the "smallest" solution, i.e. we minimize \( \| x \|^2 \) subject to the constraint \( Ax = b \). The solution of minimum norm is \( x^* = A^T (AA^T)^{-1} b \).

1.5 Regularization

Our goal remains the same: to solve the system \( Ax = b \); however, the solution we want now is the one that minimizes \( T(x) := \| Ax - b \|^2 + \lambda \| x \|^2 \) where \( \lambda > 0 \) is a regularization parameter. The unique optimal solution is given by \( x^* = (A^T A + \lambda I)^{-1} A^T b \). It can be shown that the inverse is well-defined for all \( \lambda > 0 \).

2 Exercises

You may use Julia as a computational tool.

1. Consider the function values

\[
    f(-2) = 0, \quad f(-1) = 0, \quad f(0) = 1, \quad f(1) = 0, \quad f(2) = 0.
\]

(a) Find a straight line \( f(t) = C + Dt \) that is closest (in the least squares sense) to these values.

(b) Find the parabola \( f(t) = C + Dt + Et^2 \) that is closest (in the least squares sense) to these values. *Hint: Write down the system of equations \( Ax = b \) in three unknowns \( x = (C, D, E) \) for the parabola \( f(t) \) to go through the points.*

(c) Find the closest 4th degree polynomial for these points. What is the least squares error?

(d) Find the closest 5th degree polynomial for these points. Is the solution unique? If not, find the solution with the smallest coefficients.

2. Four points in \( \mathbb{R}^3 \) have \( x,y,z \) coordinates as follows.

\[
    a = (1,0,0), \quad b = (0,1,1), \quad c = (-1,0,3), \quad d = (0,-1,4)
\]

(a) The equations \( z = C + Dx + Ey \) at points \( a, b, c, d \) are \( Ax = b \) with three unknowns \( x = (C, D, E) \). What is \( A \)?

(b) Find the plane \( z = C + Dx + Ey \) that gives the best fit to the four points \( a, b, c, d \).

(c) What is the least squares error? Check that it is equal to the squared magnitude of \( (b - Pb) \), where \( P \) is the projection matrix to the column space.

(d) Predict the value of \( z \) when \( (x, y) = (2, -1) \).

3. Tim has a large number of data points \((x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\). He plots the \( N \) data points on the \( xy \)-plane and finds out that they do not look linear but rather are exponential – that is \( y_i \approx ab^{xi} \) for some constants \( a, b \). Describe a procedure to help Tim determine a reasonable choice of \( a, b \) given the values of \( x_i \)'s and \( y_i \)'s.