Lectures on Multivariable Mathematics: Inner Products, Norms, and Orthogonalization

Dr. Chuck Rocca
roccac@wcsu.edu

Table of Contents

Objectives

After this lesson you should be able to:

1. Define inner products,

Objectives

After this lesson you should be able to:

1. Define inner products,
2. Understand the relation of inner products to norms,

Objectives

After this lesson you should be able to:

1. Define inner products,
2. Understand the relation of inner products to norms,
3. Recognize inner products beyond matrices and vectors,

Objectives

After this lesson you should be able to:

1. Define inner products,
2. Understand the relation of inner products to norms,
3. Recognize inner products beyond matrices and vectors,
4. Construct and utilize orthonormal bases,

Objectives

After this lesson you should be able to:

1. Define inner products,
2. Understand the relation of inner products to norms,
3. Recognize inner products beyond matrices and vectors,
4. Construct and utilize orthonormal bases,
5. Find the QR-Decomposition of a matrix, and

Objectives

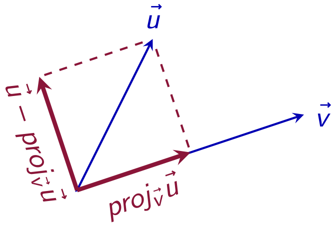
After this lesson you should be able to:

1. Define inner products,
2. Understand the relation of inner products to norms,
3. Recognize inner products beyond matrices and vectors,
4. Construct and utilize orthonormal bases,
5. Find the QR-Decomposition of a matrix, and
6. Use orthogonalization and projections to fit data to a curve.

Table of Contents

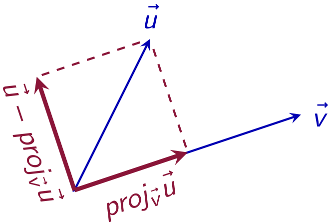
Recall Vector Projections

$$proj_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v} = \left(\frac{||\vec{u}||\,||\vec{v}||\cos(\theta)}{||\vec{v}||}\right)\left(\frac{\vec{v}}{||\vec{v}||}\right)$$



Recall Vector Projections

$$proj_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v} = \left(\frac{||\vec{u}||\,||\vec{v}||\cos(\theta)}{||\vec{v}||}\right)\left(\frac{\vec{v}}{||\vec{v}||}\right)$$



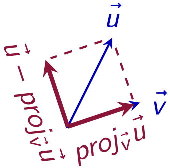
Note that we divide the $\vec{v}$ by its own length to create a vector of length one, this is called normalizing.

Orthogonal and Orthonormal Vectors

Definition (Normalizing and Normal Vectors)

Given a vector $\vec{v}$ we say that $\vec{v}/||\vec{v}||$ is the normalized version of $\vec{v}$ and is called a normal vector.

$$proj_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v} = \cos(\theta)\vec{v}$$



Orthogonal Matrices

In this product, $A^T A$, the columns of A are orthogonal so we get a diagonal matrix.

$$\begin{pmatrix} 1 & -3 & 0 \\ 3 & 1 & 1 \\ -3 & -1 & 10 \end{pmatrix} \begin{pmatrix} 1 & 3 & -3 \\ -3 & 1 & -1 \\ 0 & 1 & 10 \end{pmatrix} = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 110 \end{pmatrix}$$

Orthogonal Matrices

In this product, $A^T A$, the columns of A are orthogonal so we get a diagonal matrix.

$$\begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} & 0 \\ \frac{3}{\sqrt{11}} & \frac{1}{\sqrt{11}} & \frac{1}{\sqrt{11}} \\ -\frac{3}{\sqrt{110}} & -\frac{1}{\sqrt{110}} & \frac{10}{\sqrt{110}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{11}} & -\frac{3}{\sqrt{110}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{11}} & -\frac{1}{\sqrt{110}} \\ 0 & \frac{1}{\sqrt{11}} & \frac{10}{\sqrt{110}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If we normalize the columns, then A and A^T are inverses. These are orthogonal matrices.

Orthogonal Matrices

In this product, $A^T A$, the columns of A are orthogonal so we get a diagonal matrix.

$$\begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{11}} & -\frac{3}{\sqrt{110}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{11}} & -\frac{1}{\sqrt{110}} \\ 0 & \frac{1}{\sqrt{11}} & \frac{10}{\sqrt{110}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} & 0 \\ \frac{3}{\sqrt{11}} & \frac{1}{\sqrt{11}} & \frac{1}{\sqrt{11}} \\ -\frac{3}{\sqrt{110}} & -\frac{1}{\sqrt{110}} & \frac{10}{\sqrt{110}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

If we normalize the columns, then A and A^T are inverses. These are orthogonal matrices.

Orthogonal and Orthonormal Basis

Definition

A basis for a vector space is an orthogonal basis if all the vectors in the basis are mutually perpendicular. If, in addition, the vectors have unit length, then we say the basis is orthonormal.

$$B = \left\{ \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ -1 \\ 10 \end{pmatrix} \right\}$$

Orthogonal and Orthonormal Basis

Definition

A basis for a vector space is an orthogonal basis if all the vectors in the basis are mutually perpendicular. If, in addition, the vectors have unit length, then we say the basis is orthonormal.

$$B = \left\{ \begin{pmatrix} \frac{1}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{pmatrix} \begin{pmatrix} -\frac{3}{\sqrt{110}} \\ -\frac{1}{\sqrt{110}} \\ \frac{10}{\sqrt{110}} \end{pmatrix} \right\}$$

Orthogonal and Orthonormal Basis

Definition

A basis for a vector space is an orthogonal basis if all the vectors in the basis are mutually perpendicular. If, in addition, the vectors have unit length, then we say the basis is orthonormal.

Definition

A matrix whose columns form an orthonormal basis is called an orthogonal matrix.

$$A = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{11}} & -\frac{3}{\sqrt{110}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{11}} & -\frac{1}{\sqrt{110}} \\ 0 & \frac{1}{\sqrt{11}} & \frac{10}{\sqrt{110}} \end{pmatrix}$$

Orthogonal Complements

Definition

Given a set of vectors W in $\mathbb{R}^n$, the orthogonal complement of W, written $W^\perp$, is the set of all vectors in $\mathbb{R}^n$ perpendicular to the vectors in W.

$$W = \left\{ \begin{pmatrix} 1 \\ -3 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \right\} \qquad W^\perp = \left\{ \begin{pmatrix} -3 \\ -1 \\ 10 \end{pmatrix} \right\}$$

Orthogonal Decomposition Theorem (Statement & Example)

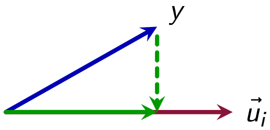
Theorem (Orthogonal Decomposition Theorem)

Let W be a subspace of $\mathbb{R}^n$, then each $y \in \mathbb{R}^n$ can be written uniquely in the form $y = \widehat{y} + z$ where $\widehat{y} \in W$ and $z \in W^{\perp}$. In fact, if $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an orthogonal basis for W, then

$$\widehat{y} = \frac{y \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{y \cdot \vec{u}_k}{\vec{u}_k \cdot \vec{u}_k}\vec{u}_k, = \sum_{\vec{u}_i} proj_{\vec{u}_i} y$$

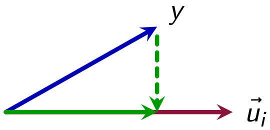
and $z = y - \widehat{y}$.

Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

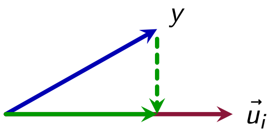
Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

$$z \cdot \vec{u}_i = (y - \widehat{y}) \cdot \vec{u}_i$$

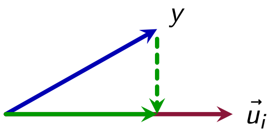
Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

$$\begin{aligned} z \cdot \vec{u}_i &= (y - \hat{y}) \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - \hat{y} \cdot \vec{u}_i \end{aligned}$$

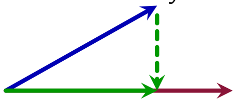
Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

$$\begin{aligned} z \cdot \vec{u}_i &= (y - \widehat{y}) \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - \widehat{y} \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - \left(\sum_{\vec{u}_i} proj_{\vec{u}_i} y \right) \cdot \vec{u}_i \end{aligned}$$

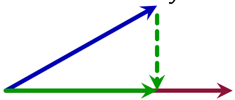
Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

$$\begin{aligned} z \cdot \vec{u}_i &= (y - \widehat{y}) \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - \widehat{y} \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - \left(\sum_{\vec{u}_i} proj_{\vec{u}_i} y \right) \cdot \vec{u}_i \\ &= y \cdot \vec{u}_i - (0 + 0 + \cdots + 0 + proj_{\vec{u}_i} y \cdot \vec{u}_i + 0 + \cdots + 0) \end{aligned}$$

Orthogonal Decomposition Theorem (Proof)



$$proj_{\vec{u}_i} y = \frac{y \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i$$

$$\begin{aligned}
z \cdot \vec{u}_i &= (y - \widehat{y}) \cdot \vec{u}_i \\
&= y \cdot \vec{u}_i - \widehat{y} \cdot \vec{u}_i \\
&= y \cdot \vec{u}_i - \left(\sum_{\vec{u}_i} proj_{\vec{u}_i} y \right) \cdot \vec{u}_i \\
&= y \cdot \vec{u}_i - (0 + 0 + \cdots + 0 + proj_{\vec{u}_i} y \cdot \vec{u}_i + 0 + \cdots + 0) \\
&= y \cdot \vec{u}_i - y \cdot \vec{u}_i = 0
\end{aligned}$$

Gram-Schmidt Process (with Normalization)

Given a basis for a vector space we can create an orthonormal basis using the Gram-Schmidt Process as follows:

1. Select one vector $\vec{u}$ from the Basis
2. Normalize $\vec{u}$ and add it to a New Basis
3. For each of the remaining vectors y in the basis
    1. Project it onto each vector $\vec{u_i}$ in the New Basis, $proj_{\vec{u}_i} y$
    2. Sum those projections, $\widehat{y} = \sum_{\vec{u}_i} proj_{\vec{u}_i} y$
    3. Subtract the sum from given vector, $z = y - \widehat{y}$
    4. Normalize this new vector, $z = z/||z||$
    5. Add the new vector to the New Basis

Gram-Schmidt Process (with Normalization)

Here we start with

$$B = \{\langle 1, 1, 0\rangle, \langle 2, 0, 1\rangle, \langle 1, 1, 1\rangle\}$$

and step through the Gram-Schmidt Process.

Gram-Schmidt Process (with Normalization)

Here we start with

$$B = \{\langle 1, 1, 0\rangle, \langle 2, 0, 1\rangle, \langle 1, 1, 1\rangle\}$$

and step through the Gram-Schmidt Process.

$$\widehat{B} = \left\{\left\langle 1/\sqrt{2}, 1/\sqrt{2}, 0\right\rangle\right\}$$

Gram-Schmidt Process (with Normalization)

Here we start with

$$B = \{\langle 1,1,0\rangle, \langle 2,0,1\rangle, \langle 1,1,1\rangle\}$$

and step through the Gram-Schmidt Process.

$$\widehat{B} = \left\{\left\langle 1/\sqrt{2}, 1/\sqrt{2}, 0\right\rangle, \left\langle 1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}\right\rangle\right\}$$

Gram-Schmidt Process (with Normalization)

Here we start with

$$B = \{\langle 1, 1, 0\rangle, \langle 2, 0, 1\rangle, \langle 1, 1, 1\rangle\}$$

and step through the Gram-Schmidt Process.

$$\widehat{B} = \left\{\left\langle 1/\sqrt{2}, 1/\sqrt{2}, 0\right\rangle, \left\langle 1/\sqrt{3}, -1/\sqrt{3}, 1/\sqrt{3}\right\rangle \left\langle -1/\sqrt{6}, 1/\sqrt{6}, 2/\sqrt{6}\right\rangle\right\}$$

Couple Important Points

Theorem (Best Approximation)

With W, y, and $\hat{y}$ as in the Orthogonal Decomposition Theorem (theorem 5), for any vector $\vec{w} \in W$,

$$||y - \hat{y}|| \leq ||y - \vec{w}||.$$

in this sense $\hat{y}$ is the nearest point to, and best approximation of, y in W.

Theorem (Orthogonal Matrices and Projections)

With W and y as in the Orthogonal Decomposition Theorem (theorem 5), if the columns of the matrix $U = [\vec{u}_1 \vec{u}_2 \cdots \vec{u}_k]$ are an orthonormal basis for W, then

$$\hat{y} = proj_W y = UU^T y.$$

QR-Decomposition: Example

Begin with A which has linearly independent columns. Apply Gram-Schmidt (with normalization) to the columns to get Q, then let $R = Q^T A$.

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad Q = \begin{pmatrix} 1/\sqrt{2} & 1\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \qquad R = Q^T A = \begin{pmatrix} \sqrt{2} & 3/\sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix}$$

QR-Decomposition: Example

Begin with A which has linearly independent columns. Apply Gram-Schmidt (with normalization) to the columns to get Q, then let $R = Q^T A$.

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad Q = \begin{pmatrix} 1/\sqrt{2} & 1\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{3} \\ 0 & 1/\sqrt{3} \end{pmatrix} \qquad R = Q^T A = \begin{pmatrix} \sqrt{2} & 3/\sqrt{2} \\ 0 & \sqrt{3} \end{pmatrix}$$

$$QR = Q(Q^T A) = (QQ^T)A = IA = A$$

QR-Decomposition: Definition

Theorem (QR Factorization)

If A is an $m \times n$ matrix with linearly independent columns and Q is an $m \times n$ matrix whose columns are an orthonormal basis for the column space of A, then there exists R an $n \times n$ upper triangular matrix with positive entries on the diagonal, specifically $R = Q^T A$.

Table of Contents

Fitting Curves to Data (Linear)

Find the equation of the line $y = mx + b$ passing through $P = (1, 2)$ and $Q = (-3, 7)$.

Fitting Curves to Data (Linear)

Find the equation of the line $y = mx + b$ passing through $P = (1, 2)$ and $Q = (-3, 7)$.

$$\begin{pmatrix} 1 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} m + b \\ -3m + b \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

Fitting Curves to Data (Linear)

Find the equation of the line $y = mx + b$ passing through $P = (1, 2)$ and $Q = (-3, 7)$.

$$\begin{pmatrix} 1 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1/4 & -1/4 \\ 3/4 & 1/4 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

Fitting Curves to Data (Linear)

Find the equation of the line $y = mx + b$ passing through $P = (1, 2)$ and $Q = (-3, 7)$.

$$\begin{pmatrix} 1 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1/4 & -1/4 \\ 3/4 & 1/4 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} -5/4 \\ 13/4 \end{pmatrix}$$

Fitting Curves to Data (Quadratic)

Find the quadratic $y = ax^2 + bx + c$ passing through (1,9), (2,15), and (-3,5).

Fitting Curves to Data (Quadratic)

Find the quadratic $y = ax^2 + bx + c$ passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & -3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -1/4 & 1/5 & 1/20 \\ -1/4 & 2/5 & -3/20 \\ 3/2 & -3/5 & 1/10 \end{pmatrix} \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

Fitting Curves to Data ("Too Much" Data)

Find a line passing through (1,9), (2,15), and (-3,5).

Fitting Curves to Data ("Too Much" Data)

Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

Fitting Curves to Data ("Too Much" Data)

Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

?

Least-Squares: Definition

Definition (Least-Squares Solution)

If A is an $m \times n$ matrix and $\vec{b}$ is in $\mathbb{R}^m$, a least-squares solution of $A\vec{x} = \vec{b}$ is an $\hat{x} \in \mathbb{R}^n$ such that

$$||\vec{b} - A\hat{x}|| \leq ||\vec{b} - A\vec{x}||$$

for all $\vec{x} \in \mathbb{R}^n$.

(Compare to the Best Approximation Theorem (thm. 6))

Least-Squares: Construction Theorem

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. *The equation $A\vec{x} = \vec{b}$ has a unique least squares solution for each $\vec{b}$.*
2. *The columns of A are linearly independent.*
3. *The matrix $A^T A$ is invertible.*

   (Compare to Orthogonal Matrices and Projections theorem (thm. 5))

If these statements are true then we can write $A = QR$ (thm. 8) and then

$$\hat{x} = (A^T A)^{-1} A^T b = R^{-1} Q^T b$$

is the least-squares solution.

Least-Squares: Example

Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

Least-Squares: Example

Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & -3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & -3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

Least-Squares: Example

Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 1 & 2 & -3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 1 & 2 & -3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 9 \\ 15 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 14 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 24 \\ 29 \end{pmatrix}$$

Least-Squares: Example

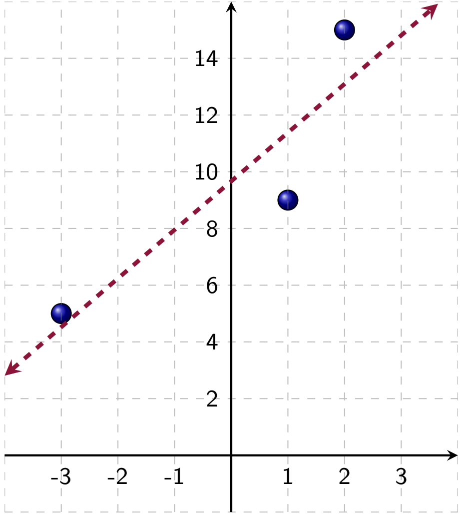
Find a line passing through (1,9), (2,15), and (-3,5).

$$\begin{pmatrix} 14 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 24 \\ 29 \end{pmatrix}$$

$$\begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 24/14 \\ 29/3 \end{pmatrix}$$

Least-Squares: Example

Find a line passing through (1,9), (2,15), and (-3,5).



Multiple Regression

Given a data set with nutritional data on 77 different cereals[1], we can fit a plane to the following linear model:

$$\mathit{calories}/\mathit{cup} = a\,(\mathit{fiber}/\mathit{cup}) + b\,(\mathit{fat}/\mathit{cup}) + c$$

by creating a matrix A whose first columns are the measures of the fiber and the fat and the third column is ones (for the constant term):

$$A = \begin{pmatrix} \mathit{fiber} & \mathit{fat} & 1 \end{pmatrix}.$$

Then multiplying $(A^T A)^{-1} A^T$ by the vector of calories we get $a \approx 3.08$, $b \approx 18.67$, and $c = 105.45$, i.e.

$$\mathit{calories}/\mathit{cup} \approx 3.08\,\mathit{fiber}/\mathit{cup} + 18.67\,\mathit{fat}/\mathit{cup} + 105.45.$$

[1] (https://domo-support.domo.com/s/article/360043931814?language=en_US)

Multiple Regression (Code)

```
# Input
import numpy as np
from numpy import linalg as la
import pandas as pd
df = pd.read_csv('cereal.csv')
cal = np.array(df['calories'])
fiber =np.array(df['fiber'])
fat = np.array(df['fat'])
cups = np.array(df['cups'])
cnt, = cal.shape
A = np.column_stack([fiber/cups,fat/cups,np.ones(cnt)])
la.inv(A.transpose()@A)@A.transpose()@(cal/cups)
# Output
array([  3.0792079 ,  18.67123465, 105.45361462])
```

Table of Contents

Inner Product: The Dot Product Generalized

Definition (Inner Product)

An inner product on a vector space V is a binary function from V to $\mathbb{R}$ such that

1. $\langle u, v \rangle = \langle v, u \rangle$
2. $\langle u + w, v \rangle = \langle u, v \rangle + \langle w, v \rangle$
3. $\langle cu, v \rangle = \langle u, cv \rangle = c\langle u, v \rangle$
4. $\langle u, u \rangle \geq 0$ with equality only when $u = 0$

A vector space with an inner product is called an inner product space.

The Norm Generalized

Definition (Norm)

In an inner product space we can define the norm of a vector as

$$||u|| = \sqrt{\langle u, u \rangle}.$$

Then distance and orthogonality are measured as before.

Weighted Inner Product in $\mathbb{R}^3$

Here is an inner product that weights the x-coordinate twice the y and the y-coordinate twice the z

$$\langle u, v \rangle = u^t A v = \begin{pmatrix} x_0 & y_0 & z_0 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = 4x_0x_1 + 2y_0y_1 + z_0z_1.$$

Since the matrix in the middle is symmetric, $A = A^T$, this will satisfy the definition of an inner product.

Inner Product for $\mathbb{P}_2$

Given two polynomials in $\mathbb{P}_2$ (quadratic and below) define

$$\langle p, q \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1).$$

Since quadratics are uniquely determined by three points, this will define an inner product on $\mathbb{P}_2$. Therefore we can make sense of norms, projections, and orthogonalization for polynomials.

Inner Product for $\mathbb{P}_2$

We can project $p(t) = 7 - t^4$ onto $\mathbb{P}_2$ with basis

$$p_0 = 1,\ p_1 = t,\ p_2 = t^2 - 2/3$$

using the previous inner product:

$$\begin{aligned} proj_{\mathbb{P}_2} p(t) &= \frac{\langle p_0, p\rangle}{\langle p_0, p_0\rangle} p_0 + \frac{\langle p_1, p\rangle}{\langle p_1, p_1\rangle} p_0 + \frac{\langle p_2, p\rangle}{\langle p_2, p_2\rangle} p_2 \\ &= \frac{19}{3} + \frac{0}{2} t - (t^2 - 2/3) \\ &= 7 - t^2. \end{aligned}$$

So the best quadratic approximation of $p(t) = 7 - t^4$ is $7 - t^2$

Inner Product with Continuous Functions

Given functions $f(x)$ and $g(x)$ continuous on $[a, b]$, i.e. $f, g \in C[a, b]$, we can define:

$$\langle f, g \rangle = \int_a^b f(x)g(x)\ dx.$$

In particular $p_0(t) = 1$ and $p_1(t) = t$ are orthogonal in $C[-1, 1]$:

$$\langle 1, t \rangle = \int_{-1}^1 t\ dx = 0.$$

And if we apply the Orthogonal Decomposition Theorem (thm. 5) to $p_2(t) = t^2$ we get:

$$z_2(t) = t^2 - \frac{\int_{-1}^1 1 \cdot t^2\ dt}{\int_{-1}^1 1 \cdot 1\ dt} + \frac{\int_{-1}^1 t \cdot t^2\ dt}{\int_{-1}^1 t \cdot t\ dt} t = t^2 - 1/3$$

which is orthogonal to the other two functions.

Cauchy-Schwarz Inequality

Theorem (Cauchy-Schwarz Inequality)

Given an inner product space V,

$$|\langle u, v\rangle| \leq ||u|| \, ||v||.$$

Cauchy-Schwarz Inequality

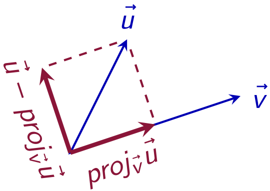
Theorem (Cauchy-Schwarz Inequality)

Given an inner product space V,

$$|\langle u, v \rangle| \leq ||u|| \, ||v||.$$

$$||proj_v u|| = \frac{|\langle u, v \rangle|}{||v||^2} ||v|| = \frac{|\langle u, v \rangle|}{||v||}$$

$$proj_{\vec{v}} \vec{u} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v}$$



Cauchy-Schwarz Inequality

Theorem (Cauchy-Schwarz Inequality)

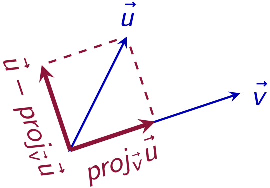
Given an inner product space V,

$$|\langle u, v\rangle| \leq ||u||\,||v||.$$

$$proj_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}$$

$$||proj_v u|| = \frac{|\langle u, v\rangle|}{||v||^2}||v|| = \frac{|\langle u, v\rangle|}{||v||}$$

$$|\langle u, v\rangle| = ||proj_v u||\,||v||$$



Cauchy-Schwarz Inequality

Theorem (Cauchy-Schwarz Inequality)

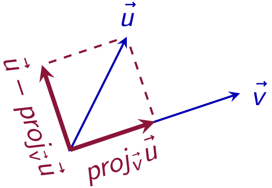
Given an inner product space V,

$$|\langle u, v\rangle| \leq ||u||\,||v||.$$

$$||proj_v u|| = \frac{|\langle u, v\rangle|}{||v||^2}||v|| = \frac{|\langle u, v\rangle|}{||v||}$$

$$|\langle u, v\rangle| = ||proj_v u||\,||v||$$
$$\leq ||u||\,||v||$$

$$proj_{\vec{v}}\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}$$



Triangle Inequality

Theorem (Triangle Inequality)

Given an inner product space V,

$$||u + v|| \leq ||u|| + ||v||.$$

Triangle Inequality

Theorem (Triangle Inequality)

Given an inner product space V,

$$||u + v|| \leq ||u|| + ||v||.$$

$$||u + v||^2 = \langle u + v, u + v \rangle$$

Triangle Inequality

Theorem (Triangle Inequality)

Given an inner product space V,

$$||u + v|| \leq ||u|| + ||v||.$$

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle \end{aligned}$$

Triangle Inequality

Theorem (Triangle Inequality)

Given an inner product space V,

$$||u + v|| \leq ||u|| + ||v||.$$

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle \\ &\leq ||u||^2 + 2||u||\,||v|| + ||v||^2 \qquad \text{Cauchy-Schwarz} \end{aligned}$$

Triangle Inequality

Theorem (Triangle Inequality)

Given an inner product space V,

$$||u + v|| \leq ||u|| + ||v||.$$

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle \\ &\leq ||u||^2 + 2||u||\,||v|| + ||v||^2 && \text{Cauchy-Schwarz} \\ &= (||u|| + ||v||)^2 \end{aligned}$$

Table of Contents

Linear Regression: Trend Analysis (Introduction)

Definition (Trend Analysis)

Given a basis of orthogonal polynomials $\{p_0, p_1, \ldots, p_k\}$ such that $deg(p_i) = i$ and a collection of data points

$$\{(x_0, y_0), (x_1, y_1), \ldots, (x_r, y_r)\},$$

which we think of as points on the graph of some function g of degree r, the projection of g onto the basis

$$\hat{g} = c_0 p_0 + c_1 p_1 + \cdots + c_k p_k$$

is called the trend function and the c_i are the trend coefficients. Note that for each i, $c_i = \langle g, p_i \rangle / \langle p_i, p_i \rangle$ and is independent of the other c_j for $i \neq j$.

Linear Regression: Trend Analysis (Example)

Let our basis be:

$$B = \left\{p_0 = 1, p_1 = t, p_2 = t^2 - 2\right\}$$

and our data set be

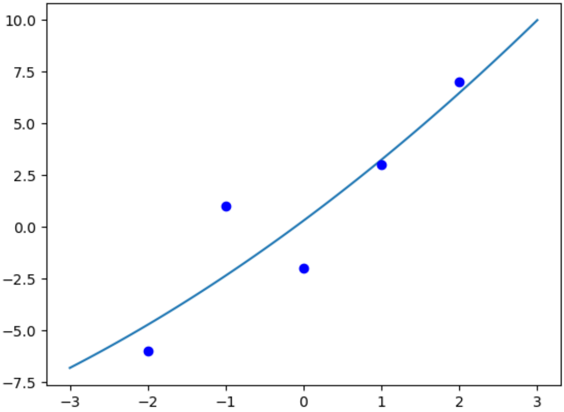
$$Data = \{(-2,-5),(-1,1),(0,-2),(1,3),(2,7)\}\,.$$

Then the quadratic that fits this data to this basis is:

$$\widehat{g} = \sum_{i=0}^{2} \frac{\langle g, p_i \rangle}{\langle p_i, p_i \rangle} p_i \approx 0.14(t^2 - 2) + 2.8t + 0.6.$$

From this we can conclude that the linear trend is more than four times the quadratic trend, so the data seems primarily linear.

Linear Regression: Trend Analysis (Example)



Fourier Analysis (Introduction)

Definition (Fourier Approximations)

A function $f(x) \in C([0, 2\pi]$ can be approximated as closely as we like by

$$\frac{a_0}{2} + a_1 \cos(t) + b_1 \sin(t) + a_2 \cos(2t) + b_2 \sin(2t) + \cdots + a_n \cos(nt) + b_n \sin(nt)$$

for sufficiently large n. This is the called a trigonometric polynomial or the n^{th}-order Fourier approximation.

Fourier Analysis (A Little Calculus)

If we define an inner product on $C[0, 2\pi]$ by $\int_0^{2\pi} f(t)g(t)\, dt$, similar to before, then the identities

$$2\cos(ax)\cos(bx) = \cos(ax + bx) + \cos(ax - bx),$$
$$2\cos(ax)\sin(bx) = \sin(ax + bx) + \sin(ax - bx), \text{ and}$$
$$2\sin(ax)\sin(bx) = \cos(ax + bx) - \cos(ax - bx),$$

can be used to show the set

$$\{1, \cos(t), \sin(t), \cos(2t), \sin(2t), \ldots, \cos(nt), \sin(nt)\}$$

form an orthogonal basis. Therefore we can find the coefficients for our Fourier approximation using projections:

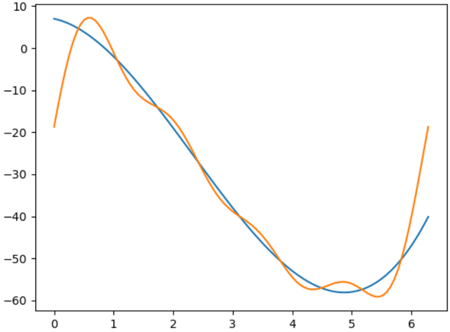
$$a_k = \frac{\langle f, \cos(kt)\rangle}{\langle \cos(kt), \cos(kt)\rangle} \text{ and } b_k = \frac{\langle f, \sin(kt)\rangle}{\langle \sin(kt), \sin(kt)\rangle}$$

for $k \geq 1$ and $a_0 = \langle f, 1\rangle / \langle 1, 1\rangle$.

Fourier Analysis (Example)

We can approximate the function $f(t) = t^3 - 7t^2 - 3t + 7$ using this approach and we get:

$$-32.59 + 27.01\sin(t) + 9.00\sin(2t) + 5.45\sin(3t) + 3.94\sin(4t)$$
$$+ 9.70\cos(t) + 2.42\cos(2t) + 1.08\cos(3t) + 0.61\cos(4t)$$



Fourier Analysis Again (Another Example)

For a data set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, we can redefine how we find the Fourier coefficients by using sums for the inner products, such as

$$\sum_{i=1}^{n} f(x_i) \cos\left(\frac{2k\pi}{\max_j(x_j)} x_i\right) \textit{ and } \sum_{i=1}^{n} f(x_i) \sin\left(\frac{2k\pi}{\max_j(x_j)} x_i\right)$$

for the numerators in the projections. And for the denominators we use

$$\sum_{i=1}^{n} \cos\left(\frac{2k\pi}{\max_j(x_j)} x_i\right) \cos\left(\frac{2k\pi}{\max_j(x_j)} x_i\right)$$

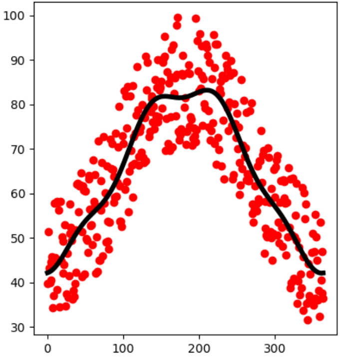
and

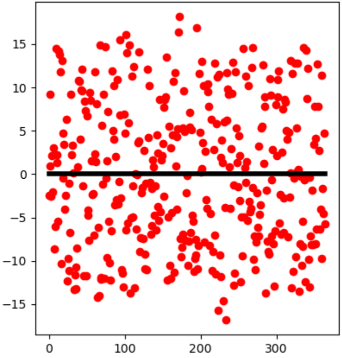
$$\sum_{i=1}^{n} \sin\left(\frac{2k\pi}{\max_j(x_j)} x_i\right) \sin\left(\frac{2k\pi}{\max_j(x_j)} x_i\right).$$

(Note this is a very primitive approach; a proper course specifically on data analysis would offer more detail.)

Fourier Analysis Again (Another Example)

If we have average daily temperature data throughout the year we can fit a curve to it:





Lectures on Multivariable Mathematics: Inner Products, Norms, and Orthogonalization

Dr. Chuck Rocca
roccac@wcsu.edu