

Ch4 Lecture 2

More Least Squares

Normal equations for multiple predictors

When we have multiple predictors, \mathbf{A} is a matrix

...

Each column corresponds to a predictor variable.

...

\mathbf{x} is a column vector of the coefficients we are trying to find, one per predictor.

...

What do the normal equations tell us? We can break them apart into one equation corresponding to each column of \mathbf{A} .

...

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$$

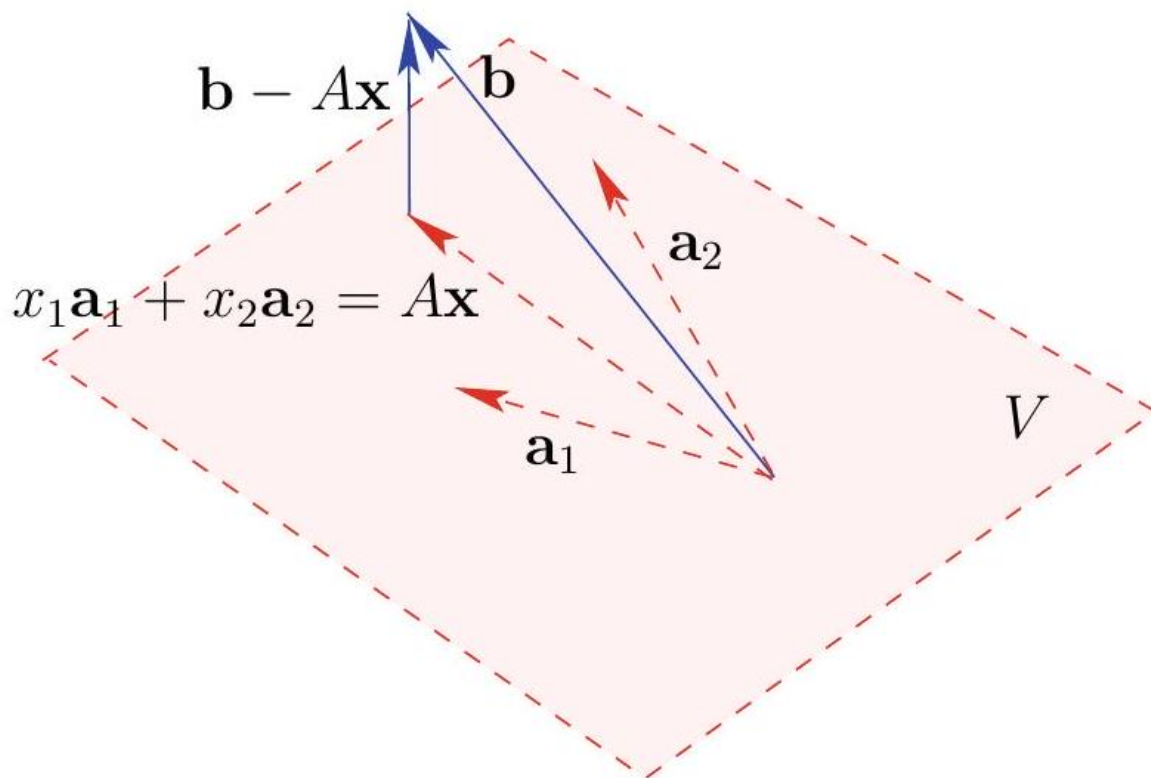
becomes

$$\mathbf{a}_i \cdot \mathbf{A} \mathbf{x} = \mathbf{a}_i \cdot \mathbf{b} \text{ for every } i$$

$$\mathbf{a}_i \cdot (\mathbf{b} - \mathbf{A} \mathbf{x}) = 0 \text{ for every } i$$

...

We need to find \mathbf{x} such that the residuals are orthogonal to every column of \mathbf{A} .



Example

We have two predictors a_1 and a_2 and a response variable b . We have the following data:

x_1	x_2	y
2	1	0
1	1	0
2	1	2

...

We are hoping to find a linear relationship of the form $\beta_1 x_1 + \beta_2 x_2 = y$ for some values of β_1 and β_2 .

...

This would bring us the following system of equations:

$$\begin{aligned}2\beta_1 + \beta_2 &= 0 \\ \beta_1 + \beta_2 &= 0 \\ 2\beta_1 + \beta_2 &= 2.\end{aligned}$$

. . .

Obviously inconsistent! ($2\beta_1 + \beta_2$ cannot equal both 0 and 2.)

. . .

Find the least squares solution using the normal equations:

Change variable names, so the β s are now \mathbf{x} , the x s are now the columns \mathbf{A} , and the y s are now \mathbf{b} .

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

. . .

Residuals are $\mathbf{b} - \mathbf{Ax}$.

Normal equations are $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$.

$$A^T A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 5 \\ 5 & 3 \end{bmatrix}$$

with inverse

$$(A^T A)^{-1} = \begin{bmatrix} 9 & 5 \\ 5 & 3 \end{bmatrix}^{-1} = \frac{1}{2} \begin{bmatrix} 3 & -5 \\ -5 & 9 \end{bmatrix}$$

. . .

$$A^T \mathbf{b} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Least Squares Solution

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b} = \frac{1}{2} \begin{bmatrix} 3 & -5 \\ -5 & 9 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

...

Back in our original variables, this means the best fit estimate for β_1 is 1 and the best β_2 is -1.

Orthogonal and Orthonormal Sets

The set of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is an **orthogonal set** if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$.

If, in addition, each vector has unit length, i.e., $\mathbf{v}_i \cdot \mathbf{v}_i = 1$, then the set of vectors is **orthonormal**.

Orthogonal Coordinates Theorem

If $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are nonzero and orthogonal, and

$$\mathbf{v} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\},$$

\mathbf{v} can be expressed uniquely (up to order) as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, namely

$$\mathbf{v} = \frac{\mathbf{v}_1 \cdot \mathbf{v}}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{v}}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 + \dots + \frac{\mathbf{v}_n \cdot \mathbf{v}}{\mathbf{v}_n \cdot \mathbf{v}_n} \mathbf{v}_n$$

Proof

Since $\mathbf{v} \in \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$,

we can write \mathbf{v} as a linear combination of the \mathbf{v}_i 's:

$$\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_n \mathbf{v}_n$$

...

Now take the inner product of both sides with \mathbf{v}_k :

...

Since $\mathbf{v}_k \cdot \mathbf{v}_j = 0$ if $j \neq k$,

$$\begin{aligned}\mathbf{v}_k \cdot \mathbf{v} &= \mathbf{v}_k \cdot (c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_n \mathbf{v}_n) \\ &= c_1 \mathbf{v}_k \cdot \mathbf{v}_1 + c_2 \mathbf{v}_k \cdot \mathbf{v}_2 + \dots + c_n \mathbf{v}_k \cdot \mathbf{v}_n = c_k \mathbf{v}_k \cdot \mathbf{v}_k\end{aligned}$$

...

Since $\mathbf{v}_k \neq 0$, $\|\mathbf{v}_k\|^2 = \mathbf{v}_k \cdot \mathbf{v}_k \neq 0$ so we can divide::

$$c_k = \frac{\mathbf{v}_k \cdot \mathbf{v}}{\mathbf{v}_k \cdot \mathbf{v}_k}$$

...

Any linear combination of an orthogonal set of nonzero vectors is the sum of its projections in the direction of each vector in the set.

Orthogonal matrix

Suppose $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ is an orthonormal basis of \mathbb{R}^n

...

Make these the column vectors of \mathbf{A} : $A = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$

...

Because the \mathbf{u}_i are orthonormal, $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$

...

Use this to calculate $A^T A$.

- its (i, j) th entry is $\mathbf{u}_n^T \mathbf{u}_n$
- $= \delta_{ij}$
- So we have $A^T A = [\delta_{ij}] = I$
- Therefore, $A^T = A^{-1}$

Orthogonal matrix

A square real matrix Q is called **orthogonal** if $Q^T = Q^{-1}$.

A square matrix U is called **unitary** if $U^* = U^{-1}$.

Example

Show that the matrix $R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ is orthogonal.

...

$$\begin{aligned} R(\theta)^T R(\theta) &= \left(\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \right)^T \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} \cos \theta \sin \theta & \\ -\sin \theta \cos \theta & \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2 \theta + \sin^2 \theta & \cos \theta \sin \theta - \sin \theta \cos \theta \\ -\cos \theta \sin \theta + \sin \theta \cos \theta & \sin^2 \theta + \cos^2 \theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

Rigidity of orthogonal (and unitary) matrices

This rotation matrix preserves vector lengths and angles between vectors (illustrate by hand).

Thus, $R(\theta)\mathbf{x} \cdot R(\theta)\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$

...

This is true in *general* for orthogonal matrices:

$$Q\mathbf{x} \cdot Q\mathbf{y} = (Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

also,

$$\|Q\mathbf{x}\|^2 = Q\mathbf{x} \cdot Q\mathbf{x} = (Q\mathbf{x})^T Q\mathbf{x} = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|^2$$

Finding orthogonal bases

Gram-Schmidt Algorithm

Let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$ be linearly independent vectors in a standard space.

...

Define vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ **recursively**:

Take each \mathbf{w}_k and subtract off the projection of \mathbf{w}_k onto each of the previous vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k-1}$

...

$$\mathbf{v}_k = \mathbf{w}_k - \frac{\mathbf{v}_1 \cdot \mathbf{w}_k}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_k}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 - \cdots - \frac{\mathbf{v}_{k-1} \cdot \mathbf{w}_k}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}} \mathbf{v}_{k-1}, \quad k = 1, \dots, n$$

$$\mathbf{v}_k = \mathbf{w}_k - \frac{\mathbf{v}_1 \cdot \mathbf{w}_k}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_k}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 - \cdots - \frac{\mathbf{v}_{k-1} \cdot \mathbf{w}_k}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}} \mathbf{v}_{k-1}, \quad k = 1, \dots, n$$

...

Then

- (1) The vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ form an orthogonal set.
- (2) For each index $k = 1, \dots, n$,

...

$$\text{span}\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$$

Example

Let $V = \mathcal{C}(A)$ with the standard inner product and compute an orthonormal basis of V , where

$$A = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 1 & -1 & 3 & 2 \\ 1 & -1 & 3 & 2 \\ -1 & 1 & -3 & 1 \end{bmatrix}$$

...

RREF of A is:

$$R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

...

The independent columns are columns 1, 2, and 4.

...

So let these be our $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$.

Step 1: $\mathbf{v}_1 = \mathbf{w}_1 = (1, 1, 1, -1)$

. . .

Step 2: $\mathbf{v}_2 = \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1$. . .

$$= (2, -1, -1, 1) - \frac{-1}{4}(1, 1, 1, -1) = \frac{1}{4}(9, -3, -3, 3)$$

. . .

Step 3: $\mathbf{v}_3 = \mathbf{w}_3 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2$

. . .

$$= (-1, 2, 2, 1) - \frac{2}{4}(1, 1, 1, -1) - \frac{-18}{108}(9, -3, -3, 3) = (0, 1, 1, 2)$$

Finding orthonormal basis

Now we need to normalize these vectors to get an orthonormal basis.

. . .

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|} \mathbf{v}_1 = \frac{1}{2}(1, 1, 1, -1)$$

. . .

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|} \mathbf{v}_2 = \frac{1}{\sqrt{108}}(9, -3, -3, 3)$$

etc.

QR Factorization

If A is an $m \times n$ full-column-rank matrix, then $A = QR$, where the columns of the $m \times n$ matrix Q are orthonormal vectors and the $n \times n$ matrix R is upper triangular with nonzero diagonal entries.

. . .

Why do we care?

- Can use to solve linear systems:

. . .

- can be less susceptible to round-off error than Gauss-Jordan.
- Can be used to solve least squares problems (will show you how!)

How to find Q and R

1. Start with the columns of A, $A = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3]$. (For now assume they are linearly independent.)
2. Do Gram-Schmidt on the columns of A:

...

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{w}_1 \\ \mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2.\end{aligned}$$

...

3. Solve these equations for $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$:

...

$$\begin{aligned}\mathbf{w}_1 &= \mathbf{v}_1 \\ \mathbf{w}_2 &= \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \mathbf{v}_2 \\ \mathbf{w}_3 &= \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 + \mathbf{v}_3.\end{aligned}$$

$$\begin{aligned}\mathbf{w}_1 &= \mathbf{v}_1 \\ \mathbf{w}_2 &= \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \mathbf{v}_2 \\ \mathbf{w}_3 &= \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 + \mathbf{v}_3.\end{aligned}$$

In matrix form, these become:

...

$$A = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3] = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] \begin{bmatrix} 1 & \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} & \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \\ 0 & 1 & \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \\ 0 & 0 & 1 \end{bmatrix}$$

Normalizing Q

We can normalize the columns of Q by dividing each column by its length.

...

Set $\mathbf{q}_j = \mathbf{v}_j / \|\mathbf{v}_j\|$.

...

$$A = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & 0 & 0 \\ 0 & \|\mathbf{v}_2\| & 0 \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix} \begin{bmatrix} 1 & \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} & \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \\ 0 & 1 & \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \\ 0 & 0 & 1 \end{bmatrix}$$

...

$$= [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\|\mathbf{v}_1\|} & \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\|\mathbf{v}_1\|} \\ 0 & \|\mathbf{v}_2\| & \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\|\mathbf{v}_2\|} \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix}.$$

Final form

$$A = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\|\mathbf{v}_1\|} & \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\|\mathbf{v}_1\|} \\ 0 & \|\mathbf{v}_2\| & \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\|\mathbf{v}_2\|} \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix}$$

$$= [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] \begin{bmatrix} \|\mathbf{v}_1\| & \mathbf{q}_1 \cdot \mathbf{w}_2 & \mathbf{q}_1 \cdot \mathbf{w}_3 \\ 0 & \|\mathbf{v}_2\| & \mathbf{q}_2 \cdot \mathbf{w}_3 \\ 0 & 0 & \|\mathbf{v}_3\| \end{bmatrix}$$

Example

Find the QR factorization of the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix}$$

...

We already found the Gram-Schmidt orthogonalization of this matrix:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{w}_1 = (1, 1, 1, -1), \\
\mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_2}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \\
&= (2, -1, -1, 1) - \frac{-1}{4}(1, 1, 1, -1) = \frac{1}{4}(9, -3, -3, 3), \\
\mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{v}_1 \cdot \mathbf{w}_3}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_3}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \\
&= (-1, 2, 2, 1) - \frac{2}{4}(1, 1, 1, -1) - \frac{-18}{108}(9, -3, -3, 3) \\
&= \frac{1}{4}(-4, 8, 8, 4) - \frac{1}{4}(2, 2, 2, -2) + \frac{1}{4}(6, -2, -2, 2) = (0, 1, 1, 2).
\end{aligned}$$

...

$$\begin{aligned}
\mathbf{u}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{2}(1, 1, 1, -1), \\
\mathbf{u}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{108}}(9, -3, -3, 3) = \frac{1}{2\sqrt{3}}(3, -1, -1, 1), \\
\mathbf{u}_3 &= \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{\sqrt{6}}(0, 1, 1, 2).
\end{aligned}$$

...

The \mathbf{u} vectors we calculated before are the \mathbf{q} vectors in the QR factorization.

$$\begin{aligned}
\|\mathbf{v}_1\| &= \|(1, 1, 1, -1)\| = 2 \text{ and } \mathbf{q}_1 = \frac{1}{2}(1, 1, 1, -1) \\
\|\mathbf{v}_2\| &= \left\| \frac{1}{4}(9, -3, -3, 3) \right\| = \frac{3}{2}\sqrt{3} \text{ and } \mathbf{q}_2 = \frac{1}{2\sqrt{3}}(3, -1, -1, 1) \\
\|\mathbf{v}_3\| &= \|(0, 1, 1, 2)\| = \sqrt{6} \text{ and } \mathbf{q}_3 = \frac{1}{\sqrt{6}}(0, 1, 1, 2)
\end{aligned}$$

...

$$\begin{aligned}
\langle \mathbf{q}_1, \mathbf{w}_2 \rangle &= \frac{1}{2}(1, 1, 1, -1) \cdot (2, -1, -1, 1) = -\frac{1}{2} \\
\langle \mathbf{q}_1, \mathbf{w}_3 \rangle &= \frac{1}{2}(1, 1, 1, -1) \cdot (-1, 2, 2, 1) = 1 \\
\langle \mathbf{q}_2, \mathbf{w}_3 \rangle &= \frac{1}{2\sqrt{3}}(3, -1, -1, 1) \cdot (-1, 2, 2, 1) = -\sqrt{3}.
\end{aligned}$$

...

$$A = \begin{bmatrix} 1/2 & 3/(2\sqrt{3}) & 0 \\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6} \\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6} \\ -1/2 & 1/(2\sqrt{3}) & 2/\sqrt{6} \end{bmatrix} \begin{bmatrix} 2 & -1/2 & 1 \\ 0 & \frac{3}{2}\sqrt{3} & -\sqrt{3} \\ 0 & 0 & \sqrt{6} \end{bmatrix} = QR$$

Solving a linear system with QR

We would like to solve the system $A\mathbf{x} = \mathbf{b}$.

...

$$QR\mathbf{x} = \mathbf{b}.$$

...

Multiply both sides by Q^T :

Since Q is orthogonal, $Q^TQ = I$ and $Q^TQR\mathbf{x} = R\mathbf{x} = Q^T\mathbf{b}$.

...

Suppose we have $\mathbf{b} = (1, 1, 1, 1)$. Then we are trying to solve

$$\begin{bmatrix} 2 & -1/2 & 1 \\ 0 & \frac{3}{2}\sqrt{3} & -\sqrt{3} \\ 0 & 0 & \sqrt{6} \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1/2 & 3/(2\sqrt{3}) & 0 \\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6} \\ 1/2 & -1/(2\sqrt{3}) & 1/\sqrt{6} \\ -1/2 & 1/(2\sqrt{3}) & 2/\sqrt{6} \end{bmatrix}^T \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

```
import sympy as sp
Q = sp.Matrix([[1/2, 3/(2*sp.sqrt(3)), 0], [1/2, -1/(2*sp.sqrt(3)), 1/sp.sqrt(6)], [1/2, -1/
R = sp.Matrix([[2, -1/2, 1], [0, 3*sp.sqrt(3)/2, -sp.sqrt(3)], [0, 0, sp.sqrt(6)]])
b = sp.Matrix([1, 1, 1, 1])
print('$$\n Q^T b ='+sp.latex(Q.T*b)+'\n$$$')
print('$$\n '+sp.latex(R)+' \mathbf{x} ='+sp.latex(Q.T*b)+'\n$$$')
```

$$Q^Tb = \begin{bmatrix} 1.0 \\ \frac{\sqrt{3}}{3} \\ \frac{2\sqrt{6}}{3} \end{bmatrix}$$

$$\begin{bmatrix} 2 & -0.5 & 1 \\ 0 & \frac{3\sqrt{3}}{2} & -\sqrt{3} \\ 0 & 0 & \sqrt{6} \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1.0 \\ \frac{\sqrt{3}}{3} \\ \frac{2\sqrt{6}}{3} \end{bmatrix}$$

...

Find

$$\mathbf{x} = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Check:

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \\ 1 & -1 & 2 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Least Squares with QR factorization

What if we have a system that's not consistent?

...

pause

...

The QR factorization can be used to solve the least squares problem $A\mathbf{x} \approx \mathbf{b}$.

Haar Wavelet Transform

Image compression idea



Haaar wavelet transform

In 1D, N data points $\{x_k\}_{k=1}^N$

...

Can average terms:

$$y_k = \frac{1}{2}(x_k + x_{k-1}), \quad k \in \mathbb{Z}$$

...

At the same time, we can apply a difference filter (“unsmoothing”):

$$z_k = \frac{1}{2}(x_k - x_{k-1}), \quad k = 1, \dots, N$$

...

Idea: take our data, apply both filters to it and keep the results.

Example

$$g(t) = \frac{3}{2} + \cos\left(\frac{\pi}{4}t\right) - \frac{1}{4}\cos\left(\frac{7\pi}{4}t\right), 0 \leq t \leq 8$$

...

Sample at $t_k = k/5, k = 0, 1, \dots, 40$

...

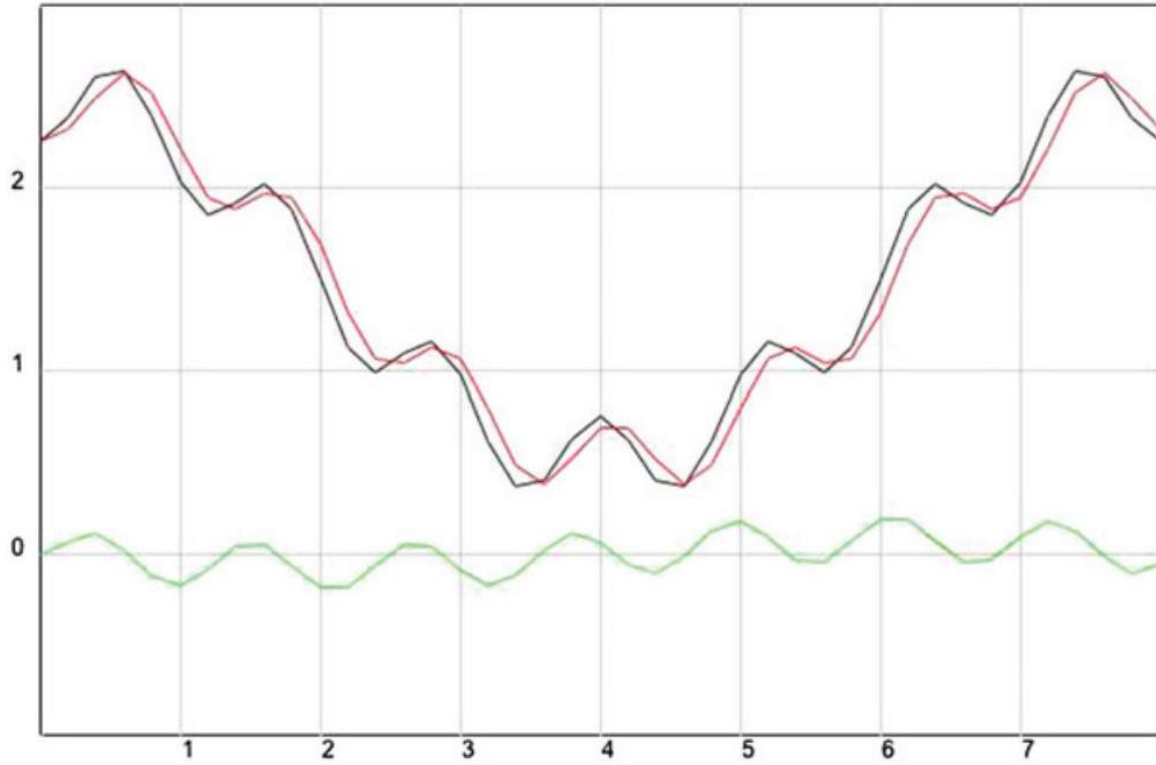


Figure 1: x: black, y: red, z: green

$$\begin{aligned} \frac{1}{2}(x_1 + x_2) &= y_2 & \frac{1}{2}(x_3 + x_4) &= y_4 & \frac{1}{2}(x_5 + x_6) &= y_6 \\ \frac{1}{2}(-x_1 + x_2) &= z_2 & \frac{1}{2}(-x_3 + x_4) &= z_4 & \frac{1}{2}(-x_5 + x_6) &= z_6 \end{aligned}$$

...

$$A_6 \mathbf{x} \equiv \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} y_2 \\ y_4 \\ y_6 \\ z_2 \\ z_4 \\ z_6 \end{bmatrix}$$

Almost orthogonal

$$A_6 A_6^T = \frac{1}{4} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$= \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{2} I_6 \quad (2)$$

In general,

$$A_N A_N^T = \frac{1}{2} I_N$$

$$A_N = \frac{1}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} U \\ \dots \\ L \end{bmatrix}$$

...

These matrices are **nearly** orthogonal.

Haar Wavelet Transform Matrix

Can make them orthogonal by dividing by $\sqrt{2}$:

$$y_k = \frac{\sqrt{2}}{2} (x_k + x_{k-1}), \quad k \in \mathbb{Z}$$

and

$$z_k = \frac{\sqrt{2}}{2} (x_k - x_{k-1}), \quad k \in \mathbb{Z}$$

...

$$W_N = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} = \frac{\sqrt{2}}{2} \begin{bmatrix} U \\ \dots \\ L \end{bmatrix}$$

Moving to 2D

For a 2D image represented pixel-by-pixel in the matrix \mathbf{A} , we can apply the 1D transform to each row and then to each column: $W_m A W_n^T$

...

Result ends up in block form:

$$W_m A W_n^T = 2 \begin{bmatrix} B & V \\ H & D \end{bmatrix}$$

...

...

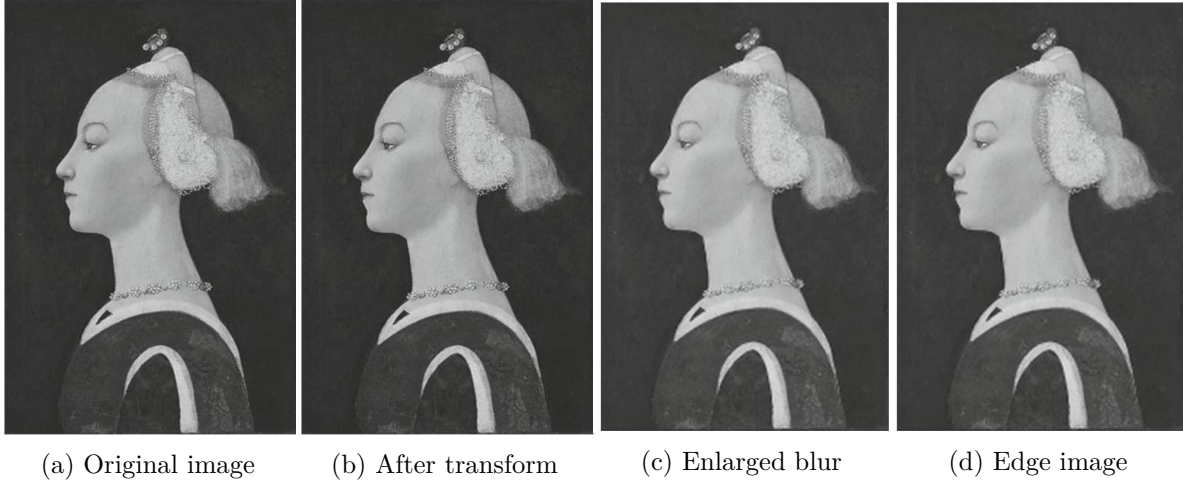


Figure 2: Effects of Haar wavelet transform on an image.

B represents the blurred image of A , while V, H and D represent edges of the image A along vertical, horizontal and diagonal directions, respectively.