Utility for solving consistent system $A\vec{x} = \vec{b}$

Utility for solving consistent system $A\vec{x} = \vec{b}$

Suppose we have the task of solving $A\vec{x} = \vec{b}$ for multiple $\vec{b}$ vectors
Suppose A is $m \times n$ with rank r

1. we compute $A = QR$ once so that
$QR\vec{x} = \vec{b}$
2. left-multiply both sides by Q^T:
$Q^TQR\vec{x} = Q^T\vec{b}$
3. from $Q^T Q = I_r$ we conclude:
$R\vec{x} = Q^T\vec{b}$ (note we do not need to invert Q)
4. computation for any new $\vec{b}$ reduces to
   - matrix-vector multiplication
   - solving an upper-triangular system

As we showed before, QR separates transformation by A into

- R: pure deformation inside Col(A)
- Q: rotation or reflection that positions that deformation in $\mathbb{R}^m$

When solving $A\vec{x} = \vec{b}$, multiplying by Q^T removes the rotational part, reducing the problem to solving only the deformation $R\vec{x} = Q^T\vec{b}$

Suppose A is rank-deficient ($r < n$)
then R is $r \times n$ and the system $R\vec{x} = Q^T\vec{b}$
represents r independent equations in n unknowns

The original system $A\vec{x} = \vec{b}$ is still solvable
if (and only if) $\vec{b}$ lies in Col(A)

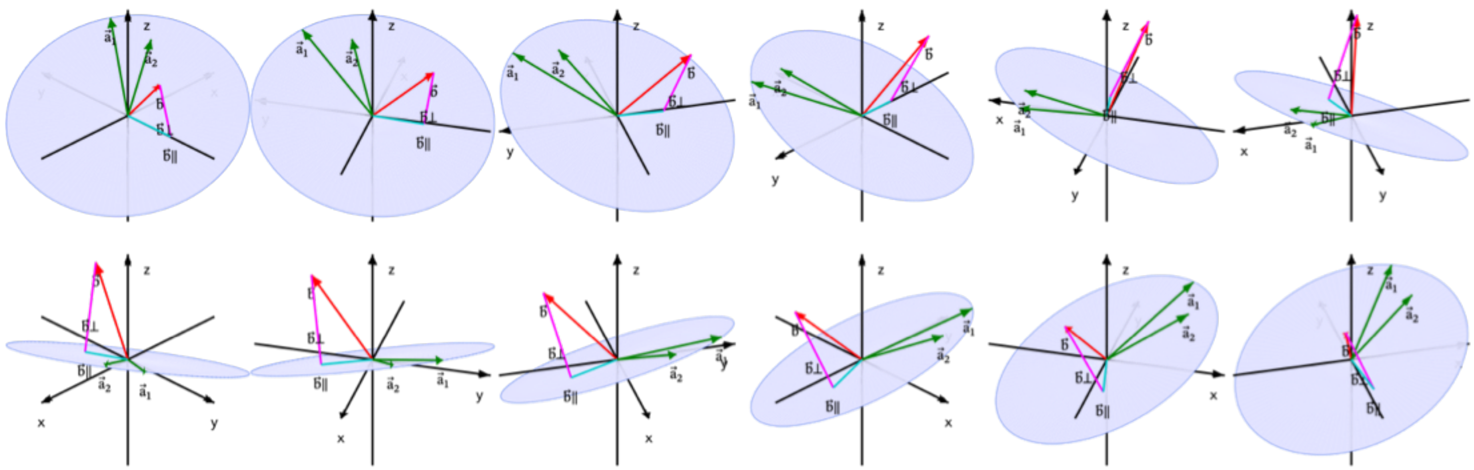
QRF vs row reduction

	REF & RREF matrix	Q matrix in QRF
Target	Rows of A	columns of A
Tools	• Subtracts multiples of rows • Scaling is not required	• Subtracts precisely chosen multiples $(a_i, q_j)\, q_j$ to eliminate directional components & achieve orthogonality • Always scales to normalize
Goal	Subtraction is done in a way that guarantees linearly independent pivot rows	Subtraction is done in a way that guarantees orthogonal columns (orthogonality ⇒ independence)
Row space	Preserves by replacing rows with linear combinations of other rows	Alters
Column space	Alters	Preserves by replacing columns with linear combinations of other columns
Solution set	Preserves	Solution set transfers to R via $R\vec{x} = Q^T\vec{b}$
Geometry	Alters	Preserves by • requiring orthogonality • normalizing
Numerical stability	May amplify rounding errors: subtracting nearly parallel rows ↓ producing very small remainders ↓ producing small pivots for later steps	While numerical errors may arise when adjacent columns are nearly parallel, errors do not affect later columns

Utility for solving an inconsistent system $A\,\vec{x} = \vec{b}$

Compare two examples of solving an inconsistent system $A\,\vec{x} = \vec{b}$ where

- $A = [\,\vec{a}_1 \mid \vec{a}_2\,]$
- $\vec{a}_1$ & $\vec{a}_2$ are vectors in $\mathbb{R}^3$, linearly independent, but not orthogonal
- $\vec{b}$ cannot be written as linear combination of $\vec{a}_1$ & $\vec{a}_2$



① As shown in the 'Four Subspaces' chapter, we can

1. rewrite the system as

$$A\,\vec{x} = \vec{b}\| + \vec{b}\perp \text{ where}$$

- $\vec{b}\perp$ is orthogonal to vectors $\vec{a}_1$ and $\vec{a}_2$
- $\vec{b}\| \in \text{col}(A)$

2. left-multiply both sides by A^T to obtain

$$A^T A\,\vec{x} = A^T\,\vec{b}\| + A^T\,\vec{b}\perp$$

3. recall that

- $\vec{b}\perp$ is orthogonal to vectors $\vec{a}_1$ and $\vec{a}_2$

$$\Leftrightarrow$$

$$\vec{b}\perp \in \ell\text{-null}(A)$$

• ℓ-null(A) is equivalent to null(A^T)

$\Leftrightarrow$

$A^T \vec{b}_\perp = \vec{0}$

4. substitute it into

$A^T A \vec{x} = A^T \vec{b}_\parallel + A^T \vec{b}_\perp = A^T \vec{b}_\parallel$

5. solve $A^T A \vec{x} = A^T \vec{b}_\parallel$ as

$\vec{x} = (A^TA)^{-1} A^T \vec{b}$

So far, this is a recap of what we have seen

Next, we solve $A^T A \vec{x} = A^T \vec{b}$ in the standard 'point of view':

$\mathbb{R}^2$ with standard canonical basis

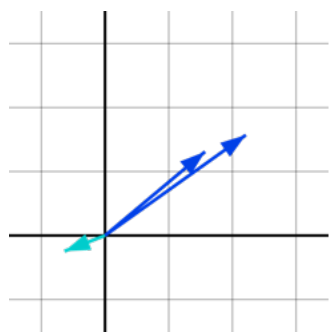
• the two columns of the 2×2 matrix $A^T A$ are drawn in blue

(note that they are not the same as vectors $\vec{a}_1$ and $\vec{a}_2$ seen in 3D)

• the target 2D vector $A^T \vec{b}$ is drawn in cyan

• solving the equation means finding coefficients x

to represent the $A^T \vec{b}$ as a linear combination of the columns of $A^T A$



② The alternative and the preferred way to solve it is to perform QR factorization on A first

Assume A has full column rank and A = Q R where

• Q has orthonormal columns, so $Q^TQ = I_r$

• R is r×r upper triangular

1. start from $A^T A \vec{x} = A^T \vec{b}$

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$

2. substitute $A = Q\,R$

3. obtain

$$A^T A = (Q\,R)^T (Q\,R) = (R^T Q^T)(Q\,R) = R^T (Q^T Q)\,R = R^T R$$

4. obtain

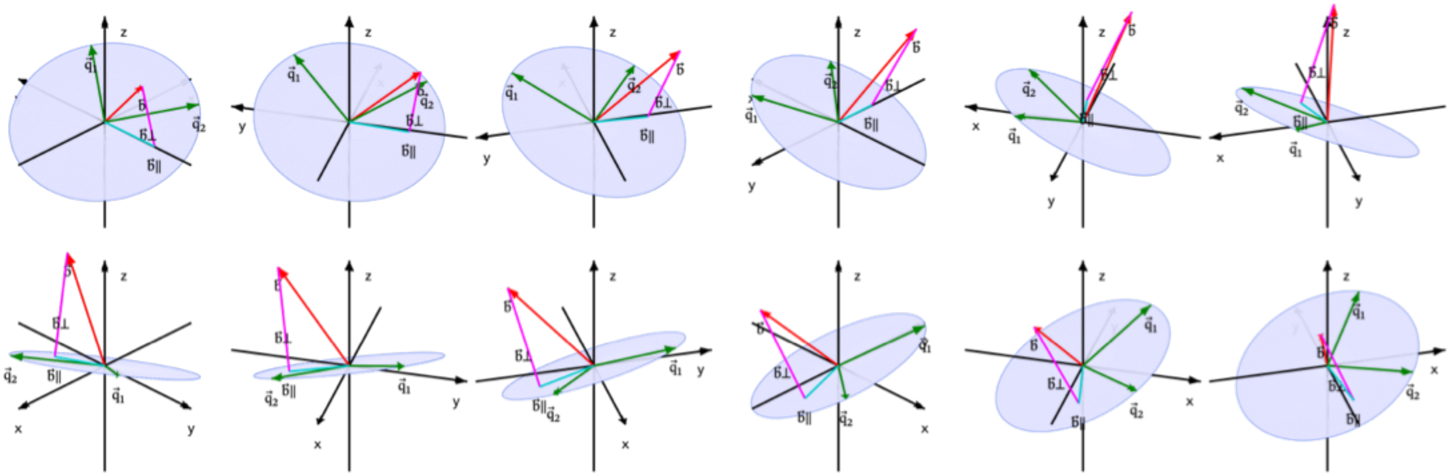
$$A^T \vec{b} = (Q\,R)^T \vec{b} = R^T Q^T \vec{b}$$

we arrive at

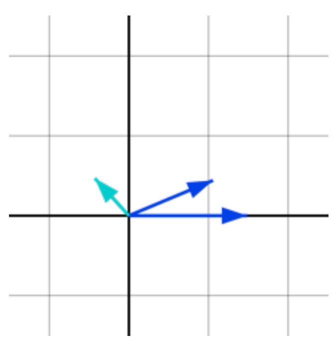
$$R\,\vec{x} = Q^T \vec{b}$$

in place of

$$A^T A\,\vec{x} = A^T \vec{b}$$



- the orthogonal columns of the 2×3 matrix Q on the 3D image are drawn in green
- the target 3D vector $\vec{b}\| = Q\,Q^T \vec{b}$ is drawn in cyan



- the columns of the upper-triangular 2×2 matrix R on the 2D image are drawn in blue
  - note that they are not the same vectors as $\vec{q}_1$ and $\vec{q}_2$
  - also note 'hierarchy' of R with first vector being aligned with $\vec{e}_1$
- the target 2D vector $Q^\intercal \vec{b}$ (different from the 3D vector $b_\parallel = Q\, Q^\intercal \vec{b}$) is drawn in cyan

Unlike the previous system,

$$R\, \vec{x} = Q^\intercal\, \vec{b}$$

is solved by back-substitution, which is computationally efficient and more numerically stable

(discussion of numerical stability is beyond the scope of this chapter)

Normal equations vs QR method

Both solve the same inconsistent equation:

$$A\, \vec{x} = \vec{b}$$

by left-multiplying both sides by $A^\intercal$:

$$A^\intercal A\, \vec{x} = A^\intercal\, \vec{b}$$

$$\left(r = \text{rank}(A)\right)$$

Normal equations	QR method
Multiplies by A^T explicitly	Factors A = Q R first to solve simplified system
Solves as $A^TA\,\vec{x} = A^T\,\vec{b}$	Solves as $R\,\vec{x} = Q^T\,\vec{b}$
Solves symmetric r×r system	Solves an upper-triangular r×r system by back-substitution
Distorts left-side by A^TA	Isolates distortion in R
Distorts right-side by A^T	Multiplies right-side by Q^T without distortion

---

OLS example 1: fitting a line

Suppose we want to fit the best line through the three points:

- (x_1 = 1.1, y_1 = -0.2)
- (x_2 = 0.5, y_2 = -0.3)
- (x_3 = 0.1, y_3 = 1.2)

Will define the line as

$y = z_1 + z_2x$

and collect the coefficients into the vector $\vec{z} = \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix}$

Substituting the three data points gives the system:

- $y_1 = z_1 + z_2 x_1$
- $y_2 = z_1 + z_2 x_2$
- $y_3 = z_1 + z_2 x_3$

or

- $-0.2 = z_1 + 1.1 z_2$
- $-0.3 = z_1 + 0.5 z_2$
- $1.2 = z_1 + 0.1 z_2$

Same system $A\,\vec{z} = [\,\vec{a}_1 \mid \vec{a}_2\,]\,\vec{z} = \vec{b}$ in matrix form:

$$\begin{bmatrix} 1 & 1.1 \\ 1 & 0.5 \\ 1 & 0.1 \end{bmatrix} \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.3 \\ 1.2 \end{bmatrix}$$

This system has

- 3 equations

$\Leftrightarrow$

data space = $\mathbb{R}^3$

- two unknowns z_1 and z_2

$\Leftrightarrow$

dimension of parameter space or column space(A) = 2

$\downarrow$

$\vec{b}$ likely lies outside parameter space

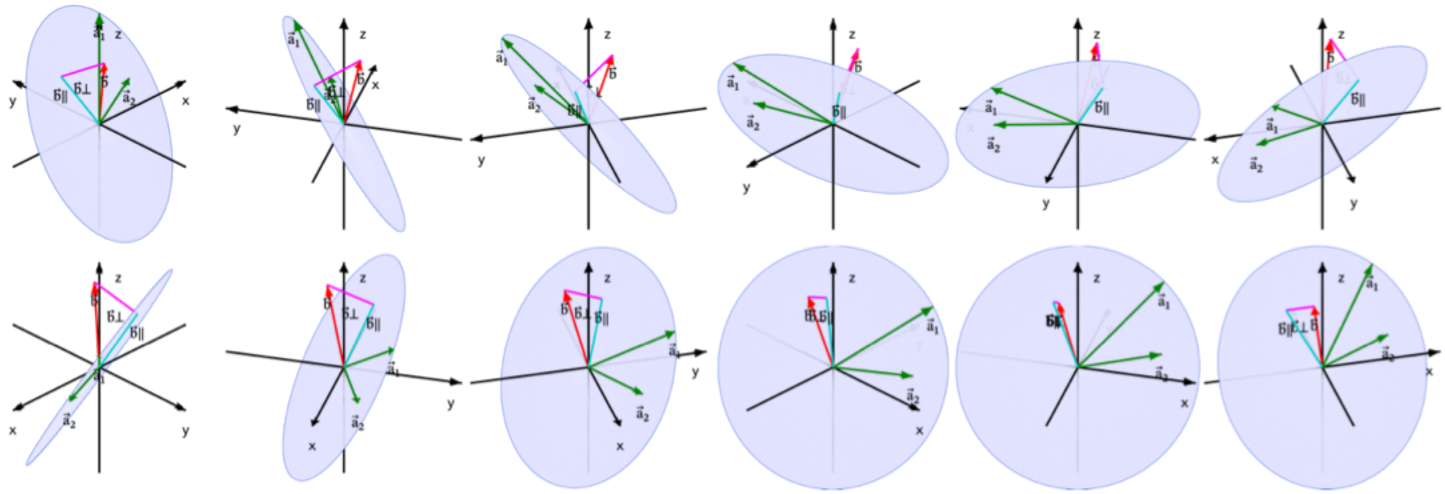
exact solution likely does not exist

The image below shows

- $\vec{a}_1$ as the green arrow corresponding to the constant term in $y = z_1 + z_2 x$
- $\vec{a}_2$ as the green arrow corresponding to the linear term in $y = z_1 + z_2 x$
- span of $\vec{a}_1$ & $\vec{a}_2$ as a plane (allowed by linear independence of $\vec{a}_1$ & $\vec{a}_2$)
- $\vec{b}$ as red arrow outside the span of $\vec{a}_1$ & $\vec{a}_2$

$\downarrow$

an exact solution indeed does not exist



One possible way to compute the least-squares solution is to solve the normal equations derived in the 'Four subspaces' chapter

$$A^\intercal A \vec{z} = A^\intercal \vec{b}$$

$$\downarrow$$

$$\vec{z} = (A^\intercal A)^{-1} A^\intercal \vec{b}$$

Since forming and inverting ($A^\intercal A$) may amplify numerical error, we orthogonalize the column space of A and solve the triangular system $R \vec{z} = Q^\intercal \vec{b}$

Converting $\vec{a}_1$ & $\vec{a}_2$ into an orthonormal basis yields

$$Q = [\, \vec{q}_1 \mid \vec{q}_2 \,] = \begin{bmatrix} \approx 0.577 & \approx 0.749 \\ \approx 0.577 & \approx -0.094 \\ \approx 0.577 & \approx -0.656 \end{bmatrix}$$

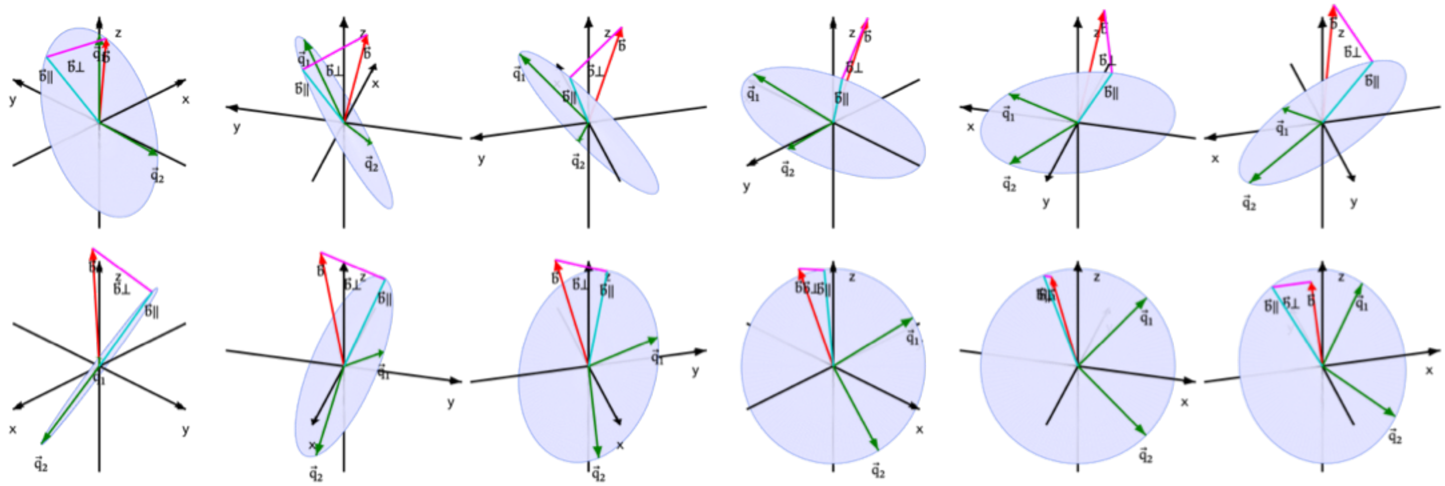
Same image with $\vec{q}_1$ & $\vec{q}_2$ in place of $\vec{a}_1$ & $\vec{a}_2$:

- $\vec{q}_1$ is same as normalized $\vec{a}_1$
- $\vec{q}_2$ is orthogonal to $\vec{q}_1$

• the plane spanned by $\vec{q}_1$ & $\vec{q}_2$ is the same as the plane spanned by $\vec{a}_1$ & $\vec{a}_2$

• $\vec{b}$ is unchanged

• the cyan vector $\vec{b}\|$ = Q Q^{T} $\vec{b}$ or projection of $\vec{b}$ onto span($\vec{q}_1$, $\vec{q}_2$)



As shown on the previous page, we compute

$$\text{upper-triangular matrix } R = Q^{T} A = \begin{bmatrix} \approx 1.732 & \approx 0.981 \\ 0 & \approx 0.712 \end{bmatrix}$$

and solve R $\vec{z}$ = Q^{T} $\vec{b}$ without inverting Q:

• z_1 ≈0.957 (constant term in $y = z_1 + z_2x$)

• z_2 ≈-1.276 (linear term in $y = z_1 + z_2x$)

• $\vec{b}$ represents the three actual 'y' values of the dataset

• Each vector in the plane span($\vec{a}_1$, $\vec{a}_2$) = span($\vec{q}_1$, $\vec{q}_2$) represents the three predicted values

of some line $y = z_1 + z_2x$ evaluated at the three sample points

• Cyan vector $\vec{b}\|$ represents the predicted values for the best fit line

• 3D magenta segment represents the residual vector

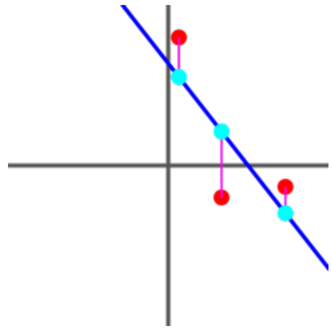
$$\vec{b}\perp = \vec{b} - \vec{b}\|$$
$$= \vec{b} - Q\,Q^{T}\,\vec{b}$$
$$= \vec{b} - A\,\vec{z}$$

• sum of squared residuals

$|\vec{b} - A\vec{z}|^2 \approx 0.581$

is the quantity the algorithm minimizes

Image below shows the best fit line $y = z_1 + z_2x$



• red dots are the 3 inputed values

• cyan dots are the 3 values predicted by the best line

• magenta lines are individual residuals or individual entries of $\vec{b}\perp$

• The computation approach worked because $\vec{a}_1$ & $\vec{a}_2$ are linearly independent and the plane exists

• To depict it as a 3D image, we are limited to

‣ 3 data points or data space = $\mathbb{R}^3$

‣ 2 parameters in $y = z_1 + z_2x$ or dimension of column space = 2

Next, we will show that the same least-squares framework can fit a nonlinear function

---

OLS example 2: fitting a parabola with 2 parameters

Suppose we want to fit a nonlinear function
through the same set of points that we used for linear regression:

- ($x_1 = 1.1$, $y_1 = -0.2$)
- ($x_2 = 0.5$, $y_2 = -0.3$)
- ($x_3 = 0.1$, $y_3 = 1.2$)

In order to show the 3D image, we are limited to 3 points and 2 parameters,
so we will define the function as a parabola with
a constant term and a quadratic term:

$$y = z_1 + z_2x^2$$

and collect the coefficients into the vector $\vec{z} = \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix}$

Substituting the three data points gives the system:

- $y_1 = z_1 + z_2x_1^2$
- $y_2 = z_1 + z_2x_2^2$
- $y_3 = z_1 + z_2x_3^2$

or

- $-0.2 = z_1 + 1.21z_2$
- $-0.3 = z_1 + 0.25z_2$
- $1.2 = z_1 + 0.01z_2$

Same system $A\,\vec{z} = [\,\vec{a}_1 \mid \vec{a}_2\,]\,\vec{z} = \vec{b}$ in matrix form:

$$\left[\begin{array}{c|c} 1 & 1.21 \\ \hline 1 & 0.25 \\ \hline 1 & 0.01 \end{array}\right] \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix} = \begin{bmatrix} -0.2 \\ -0.3 \\ 1.2 \end{bmatrix}$$

As in the previous example, this overdetermined system has

- two unknowns z_1 and z_2
- 3 equations

The image below shows

- $\vec{a}_1$ as the green arrow corresponding to the constant term in $y = z_1 + z_2x^2$

$\left(\vec{a}_1 = [1, 1, 1]^T \text{ as in previous example}\right)$

- $\vec{a}_2$ as the green arrow corresponding to the quadratic term in $y = z_1 + z_2x^2$

($\vec{a}_2$ is different from $\vec{a}_2$ in prior example)

- span of linearly independent $\vec{a}_1$ & $\vec{a}_2$ as a plane

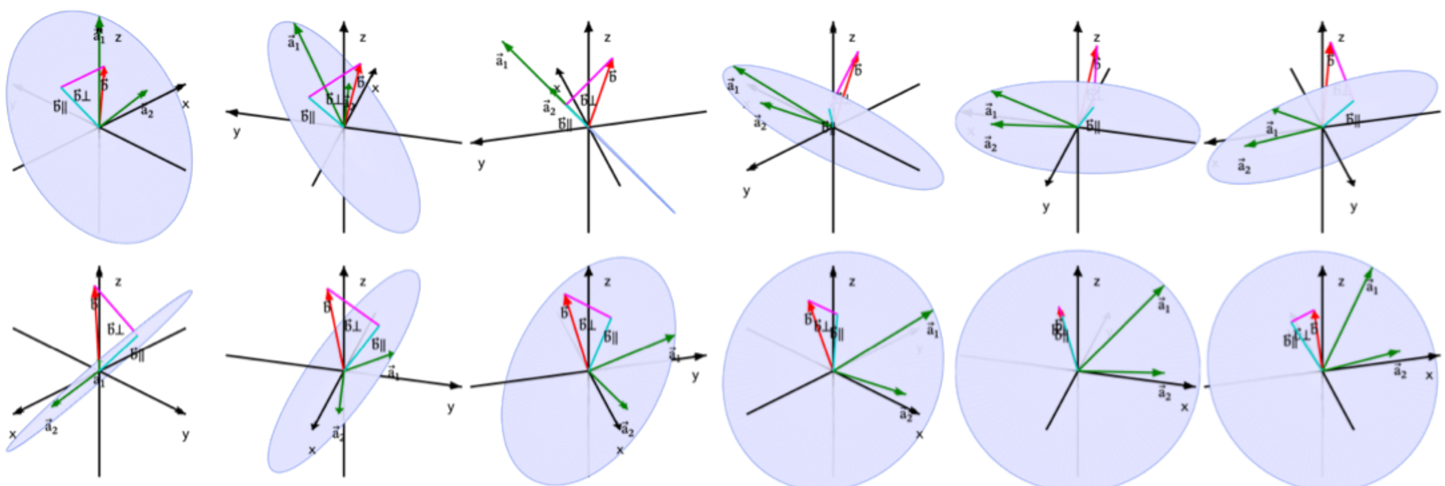
(plane is different from the plane in the prior example)

- $\vec{b}$ as red arrow (same as in prior example)

$\vec{b}$ is again outside the span of $\vec{a}_1$ & $\vec{a}_2$

↓

an exact solution does not exist

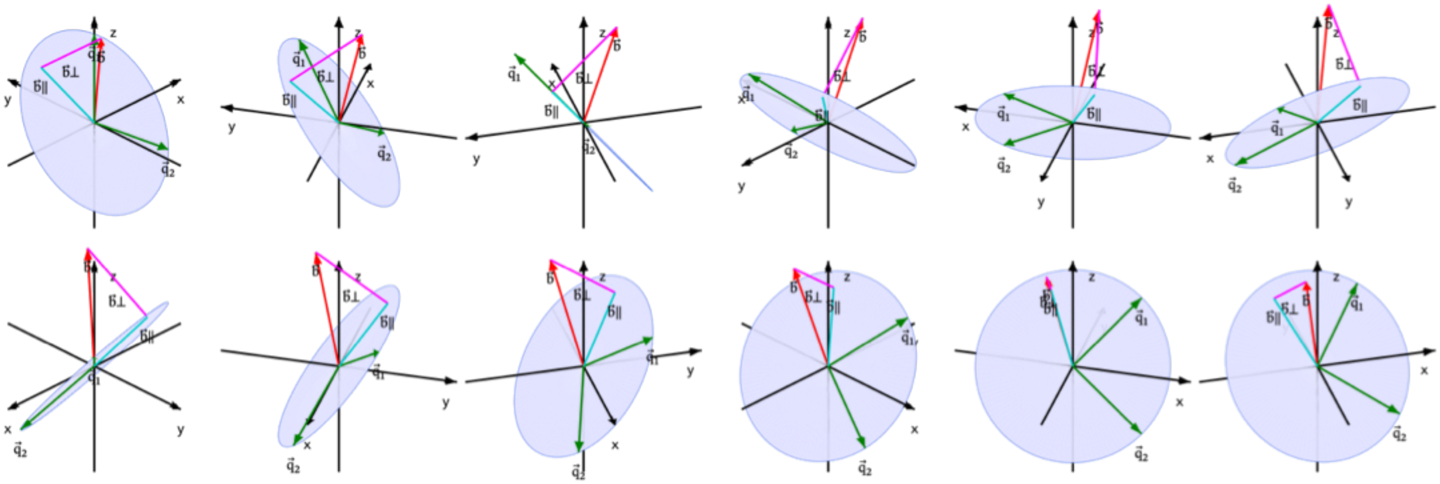


As before, we orthogonalize the column space of A:

$$Q = [\, \vec{q}_1 \mid \vec{q}_2 \,] = \begin{bmatrix} \approx 0.577 & \approx 0.802 \\ \approx 0.577 & \approx -0.267 \\ \approx 0.577 & \approx -0.535 \end{bmatrix}$$

- $\vec{q}_1$ is same as normalized $\vec{a}_1$
- the plane spanned by $\vec{q}_1$ & $\vec{q}_2$ is the same as the plane spanned by $\vec{a}_1$ & $\vec{a}_2$

• $\vec{q}_2$ is orthogonal to $\vec{q}_1$ in the same plane
• $\vec{b}$ is unchanged
• $\vec{b}_\parallel = Q\,Q^T\,\vec{b}$ or projection of $\vec{b}$ onto span($\vec{q}_1$, $\vec{q}_2$)



We compute an upper-triangular matrix $R = Q^T A = \begin{bmatrix} \approx 1.732 & \approx 0.849 \\ 0 & \approx 0.898 \end{bmatrix}$

After solving $R\,\vec{z} = Q^T\,\vec{b}$ we obtain
• $z_1 \approx 0.627$ (constant term in $y = z_1 + z_2 x^2$)
• $z_2 \approx -0.804$ (square term in $y = z_1 + z_2 x^2$)

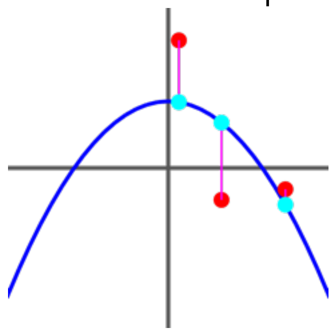
• $\vec{b}$ represents the three actual 'y' values of the dataset
• Each vector in the plane span($\vec{a}_1$, $\vec{a}_2$) = span($\vec{q}_1$, $\vec{q}_2$) represents the three predicted values
of some parabola $y = z_1 + z_2 x^2$ evaluated at the three sample points
• Cyan vector $\vec{b}_\parallel$ represents the predicted values for the best fit parabola
• 3D magenta segment represents the residual vector

$$\vec{b}_\perp = \vec{b} - \vec{b}_\parallel$$
$$= \vec{b} - Q\,Q^T\,\vec{b}$$
$$= \vec{b} - A\,\vec{z}$$

• sum of squared residuals
$| \vec{b} - A\,\vec{z} |^2 \approx 0.886$

is the quantity the algorithm minimizes

Image below shows the best fit parabola $y = z_1 + z_2x^2$



- red dots are the 3 inputed values
- cyan dots are the 3 values predicted by the best parabola
- magenta lines are individual residuals or individual entries of $\vec{b}\perp$

Two models for same data set

① Suppose we want to fit a line $y = z_1 + z_2x$ through the following point set:

- $(x_1 = 1, y_1 = 1)$
- $(x_2 = -1, y_2 = 1)$
- $(x_3 = 0, y_3 = -2)$

This gives us the following system:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

Notice that $\vec{a}_1 \cdot \vec{b} = 0$ and $\vec{a}_2 \cdot \vec{b} = 0$

↓

red vector $\vec{b}$ shown on previous pages is orthogonal to the plane

↓

$\vec{b}_{||} = \vec{0}$

$$Q = [\, \vec{q}_1 \mid \vec{q}_2 \,] = \begin{bmatrix} \approx 0.577 & \approx 0.707 \\ \approx 0.577 & \approx -0.707 \\ \approx 0.577 & 0 \end{bmatrix}$$

$\vec{q}_1$ & $\vec{q}_2$ are within span of $\vec{a}_1$ & $\vec{a}_2$

↓

$Q^T \vec{b} = \vec{0}$

$$R = Q^T A = \begin{bmatrix} \approx 1.732 & 0 \\ 0 & \approx 1.414 \end{bmatrix}$$

solving $R\,\vec{z} = Q^T \vec{b}$

$$\begin{bmatrix} \approx 1.732 & 0 \\ 0 & \approx 1.414 \end{bmatrix} \begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- $z_1 = 0$ (constant term in $y = z_1 + z_2x$)
- $z_2 = 0$ (linear term in $y = z_1 + z_2x$)

↓

data set cannot be represented by a linear model

(note that in this example, $\vec{a}_1 \cdot \vec{a}_2 = 0$, which is not a necessary condition)

② Next, we will fit a parabola $y = z_1 + z_2x^2$
through the same point set
As we saw on the previous page,
• $\vec{b}$ will remain the same
• $\vec{a}_1$ will remain the same
• $\vec{a}_2$ will change direction & the plane will change orientation

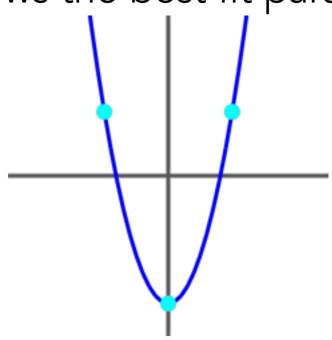
$$Q = [\,\vec{q}_1 \mid \vec{q}_2\,] = \begin{bmatrix} \approx 0.577 & \approx 0.408 \\ \approx 0.577 & \approx 0.408 \\ \approx 0.577 & \approx -0.816 \end{bmatrix}$$

solving $R\,\vec{z} = Q^{T}\,\vec{b}$

$$\begin{bmatrix} \approx 1.732 & \approx 1.155 \\ 0 & \approx 0.816 \end{bmatrix}\begin{bmatrix} \vec{z}_1 \\ \vec{z}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \approx 2.449 \end{bmatrix}$$

• $z_1 = -2$ (constant term in $y = z_1 + z_2x$)
• $z_2 = 3$ (linear term in $y = z_1 + z_2x$)

Image below shows the best fit parabola $y = z_1 + z_2x^2$



Notice that the parabola fits the data set precisely
that means that $\vec{b}\perp = \vec{0}$
↓

$\vec{b} \in \text{col}(A)$

You can also see why a straight line could not fit the same point set

In this example, replacing x with x^2 changed the plane from one orthogonal to $\vec{b}$ to one containing $\vec{b}$

Higher-dimensional set

Next, we will fit a sum of sine and cosine waves:

$$y = z_1 + z_2 \sin(x) + z_3 \cos(x) + z_4 \sin(2x) + z_5 \cos(2x)$$

into the following data set

- $(x_1 = -1.88, y_1 = 0.62)$
- $(x_2 = -1.41, y_2 = -0.17)$
- $(x_3 = -1.07, y_3 = 1.34)$
- $(x_4 = -0.73, y_4 = 0.21)$
- $(x_5 = -0.31, y_5 = -1.08)$
- $(x_6 = 0.12, y_6 = 0.57)$
- $(x_7 = 0.44, y_7 = -0.64)$
- $(x_8 = 0.91, y_8 = 1.11)$
- $(x_9 = 1.36, y_9 = 0.18)$
- $(x_{10} = 1.94, y_{10} = -0.86)$

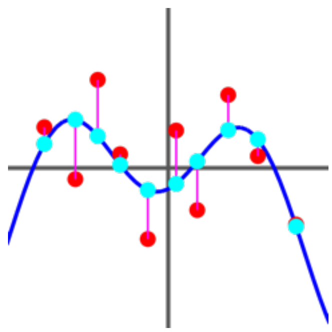
This time, we will be projecting $\vec{b}$ in $\mathbb{R}^{10}$ onto 5-dimensional col space(A)

$$A = \begin{bmatrix} 1 & \sin(x_1) & \cos(x_1) & \sin(2x_1) & \cos(2x_1) \\ 1 & \sin(x_2) & \cos(x_2) & \sin(2x_2) & \cos(2x_2) \\ 1 & \sin(x_3) & \cos(x_3) & \sin(2x_3) & \cos(2x_3) \\ 1 & \sin(x_4) & \cos(x_4) & \sin(2x_4) & \cos(2x_4) \\ 1 & \sin(x_5) & \cos(x_5) & \sin(2x_5) & \cos(2x_5) \\ 1 & \sin(x_6) & \cos(x_6) & \sin(2x_6) & \cos(2x_6) \\ 1 & \sin(x_7) & \cos(x_7) & \sin(2x_7) & \cos(2x_7) \\ 1 & \sin(x_8) & \cos(x_8) & \sin(2x_8) & \cos(2x_8) \\ 1 & \sin(x_9) & \cos(x_9) & \sin(2x_9) & \cos(2x_9) \\ 1 & \sin(x_{10}) & \cos(x_{10}) & \sin(2x_{10}) & \cos(2x_{10}) \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \end{bmatrix}$$

$$\begin{bmatrix} 1 & \approx -0.953 & \approx -0.304 & \approx 0.58 & \approx -0.815 \\ 1 & \approx -0.987 & \approx 0.16 & \approx -0.316 & \approx -0.949 \\ 1 & \approx -0.877 & \approx 0.48 & \approx -0.842 & \approx -0.539 \\ 1 & \approx -0.667 & \approx 0.745 & \approx -0.994 & \approx 0.111 \\ 1 & \approx -0.305 & \approx 0.952 & \approx -0.581 & \approx 0.814 \\ 1 & \approx 0.12 & \approx 0.993 & \approx 0.238 & \approx 0.971 \\ 1 & \approx 0.426 & \approx 0.905 & \approx 0.771 & \approx 0.637 \\ 1 & \approx 0.79 & \approx 0.614 & \approx 0.969 & \approx -0.247 \\ 1 & \approx 0.978 & \approx 0.209 & \approx 0.409 & \approx -0.912 \\ 1 & \approx 0.933 & \approx -0.361 & \approx -0.673 & \approx -0.74 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix} = \begin{bmatrix} 0.62 \\ -0.17 \\ 1.34 \\ 0.21 \\ -1.08 \\ 0.57 \\ -0.64 \\ 1.11 \\ 0.18 \\ -0.86 \end{bmatrix}$$

We solve it in the same fashion by
• orthogonalizing col space(A)
• solving upper-triangular system $R\vec{z} = Q^T\vec{b}$:

- $z_1 \approx -0.596$ (constant term)
- $z_2 \approx -0.316$ $\left(\sin(x) \text{ term}\right)$
- $z_3 \approx 1.272$ $\left(\cos(x) \text{ term}\right)$
- $z_4 \approx 0.408$ $\left(\sin(2x) \text{ term}\right)$
- $z_5 \approx -0.996$ $\left(\cos(2x) \text{ term}\right)$



Summary: Models with one input variable

① Which models correspond to linear least-squares problems
• Valid models:
parameters appear only as linear coefficients:

$$y = z_1 f_1(x) + z_2 f_2(x) + \ldots + z_n f_n(x)$$

Invalid models:

- parameters raised to power: $y = z_1 + z_2^2 x$
- **parameters used as exponents: $z_1 + e^{z2}$**
- parameters used as function arguments: $y = z_1 + \sin(z_2 x)$
- products of two or more parameters: $y = z_1 + z_2 z_3$

Valid model examples

Model	Basis functions $f_i(x)$	Utility
Constant	1	Data fluctuates around a constant level
Line	1, x	Output varies linearly with one variable
Parabola	1, x^2	Curved trend with one turning point
Polynomial	1, x, x^2, x^3, …	Smooth curve of unknown shape
Fourier function sum of sine and cosine waves	1, sin(x), cos(x), sin(2x), cos(2x)	Periodic behavior

② General form of OLS system based on general model

$$y = z_1 f_1(x) + z_2 f_2(x) + \ldots + z_n f_n(x)$$

shown for m = 5 points and n = 3 parameters

$A\vec{z} = \vec{b}$ or

$$\begin{bmatrix} f_1(x_1) & f_2(x_1) & f_3(x_1) \\ f_1(x_2) & f_2(x_2) & f_3(x_2) \\ f_1(x_3) & f_2(x_3) & f_3(x_3) \\ f_1(x_4) & f_2(x_4) & f_3(x_4) \\ f_1(x_5) & f_2(x_5) & f_3(x_5) \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

- columns(A) = basis functions
- rows(A) = data points

Assuming $m > n$ & A has full-column rank:

- $\vec{b} \in \text{col}(A) \rightarrow$ exact solution
- $\vec{b} \notin \text{col}(A) \rightarrow$ least-squares projection onto col(A)
- $\vec{b} \perp \text{col}(A) \rightarrow$ projection is zero $\rightarrow$ model cannot explain the data

④ Note that one of the f functions is usually 1 which corresponds to the constant term in
$y = z_1 f_1(x) + z_2 f_2(x) + \ldots + z_n f_n(x)$
This creates a column $[1, 1, \ldots, 1]^T$ in matrix A

$\vec{b}\perp$ is orthogonal to every column of A

$$\downarrow$$

$$[1, 1, \ldots, 1]^T \, \vec{b}\perp = \sum (b\perp)_i = 0$$

$$\Leftrightarrow$$

sum of all entries of $b\perp$ (or all residuals) is 0

$$\Leftrightarrow$$

model is centered on the data

Model with 2 input variables

Suppose we want to fit the best plane through the points:

- $(x_1 = 0.8, y_1 = 1.2, z_1 = 2.1)$
- $(x_2 = 2, y_2 = 0.7, z_2 = 3.4)$
- $(x_3 = 1.5, y_3 = 2.4, z_3 = 4)$
- $(x_4 = 3.1, y_4 = 1.8, z_4 = 5.2)$
- $(x_5 = 2.6, y_5 = 3.3, z_5 = 6.1)$

We define the plane as

$z = u_1 + u_2\, x + u_3\, y$

Solving $A\,\vec{u} = \vec{b}$ or

$$\begin{bmatrix} 1 & 0.8 & 1.2 \\ 1 & 2 & 0.7 \\ 1 & 1.5 & 2.4 \\ 1 & 3.1 & 1.8 \\ 1 & 2.6 & 3.3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 2.1 \\ 3.4 \\ 4 \\ 5.2 \\ 6.1 \end{bmatrix}$$

- $u_1 \approx 0.314$ (constant term)
- $u_2 \approx 1.164$ ('x' term)
- $u_3 \approx 0.808$ ('y' term)