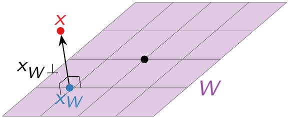
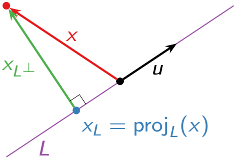
Midterm 3

Selected topics

Motivation



Example with a line: The closest point to x in L is $\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u$



Let $u = \binom{3}{2}$ and let $L = \text{Span}\{u\}$. Let $x = \binom{-6}{4}$. In this case,

$$x_L = \text{proj}_L(x) = -\frac{10}{13}\begin{pmatrix}3\\2\end{pmatrix} \qquad x_{L^\perp} = x - \text{proj}_L(x) = \begin{pmatrix}-6\\4\end{pmatrix} + \frac{10}{13}\begin{pmatrix}3\\2\end{pmatrix}.$$

Motivation

The procedures in §6 start with an *orthogonal basis* $\{u_1, u_2, \ldots, u_m\}$.

- Find the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i$$

- Find the orthogonal projection of a vector x onto the span W of $u_1, u_2, \ldots, u_m$:

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Problem: What if your basis isn't orthogonal?

Solution: The **Gram–Schmidt process**: take any basis and *make it orthogonal*.

Orthogonal Complements

Definition

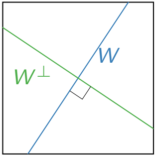
Let W be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is

$$W^\perp = \{ v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W \} \qquad \text{read “}W\text{ perp”}.$$

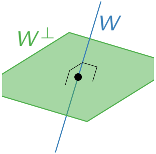
$W^\perp$ is orthogonal complement
A^T is transpose

Pictures:

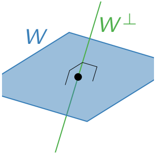
The orthogonal complement of a line in $\mathbf{R}^2$ is the perpendicular line.



The orthogonal complement of a line in $\mathbf{R}^3$ is the perpendicular plane.



The orthogonal complement of a plane in $\mathbf{R}^3$ is the perpendicular line.



Orthogonal Complements

Basic properties

Facts: Let W be a subspace of $\mathbf{R}^n$.

1. $W^\perp$ is also *a subspace of* $\mathbf{R}^n$
2. $(W^\perp)^\perp = W$
3. $\dim W + \dim W^\perp = n$
4. If $W = \text{Span}\{v_1, v_2, \ldots, v_m\}$, then

$$\begin{aligned} W^\perp &= \text{all vectors orthogonal to each } v_1, v_2, \ldots, v_m \\ &= \{x \text{ in } \mathbf{R}^n \mid x \cdot v_i = 0 \text{ for all } i = 1, 2, \ldots, m\} \\ &= \text{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}. \end{aligned}$$

Property 4

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}$$

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \text{Span}\{v_1, v_2, v_3\} = \mathbf{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

Run Gram–Schmidt:

1. $u_1 = v_1$
2. $u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
3. $u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$$= \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \frac{4}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{1}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Remember: This is an orthogonal basis for the *same subspace* W.

QR Factorization

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is *orthonormal* if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Orthonormal

A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where **Q has orthonormal columns** and *R is upper-triangular* with positive diagonal entries.

- The columns of A are a basis for $W = \operatorname{Col} A$.
- The columns of Q are equivalent *basis coming from Gram–Schmidt* (as applied to the columns of A), *after normalizing* to unit vectors.
- The columns of R *come from the steps* in Gram–Schmidt.

Find the QR factorization of $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & -2 \\ 1 & -1 & 0 \end{pmatrix}$.

(The columns are vectors from example 3.)

Step 1: Run Gram–Schmidt and *solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3*:

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_1 = u_1$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = v_2 - \frac{3}{2} u_1 = \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix} \qquad v_2 = \frac{3}{2} u_1 + u_2$$

$$u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 = v_3 + \frac{4}{5} u_2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix} \qquad v_3 = -\frac{4}{5} u_2 + u_3$$

QR Factorization

Through a second example, continued

$$v_1 = 1\, u_1 \qquad v_2 = \frac{3}{2}\, u_1 + 1\, u_2 \qquad v_3 = 0\, u_1 - \frac{4}{5}\, u_2 + 1\, u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal columns* u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

$$\widehat{Q} = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix}$$

$$\widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

QR Factorization

Through a second example, continued

$$A = \widehat{Q}\widehat{R} \qquad \widehat{Q} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix} \qquad \widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: To get Q and R, *normalize the columns* of $\widehat{Q}$ *and scale the rows* of $\widehat{R}$:

$$Q = \begin{pmatrix} | & | & | \\ u_1/\|u_1\| & u_2/\|u_2\| & u_3/\|u_3\| \\ | & | & | \end{pmatrix}$$

$$R = \begin{pmatrix} 1 \cdot \|u_1\| & 3/2 \cdot \|u_1\| & 0 \cdot \|u_1\| \\ 0 & 1 \cdot \|u_2\| & -4/5 \cdot \|u_2\| \\ 0 & 0 & 1 \cdot \|u_3\| \end{pmatrix}$$

The **final QR decomposition** is

$$A = QR \qquad Q = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix} \qquad R = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}.$$

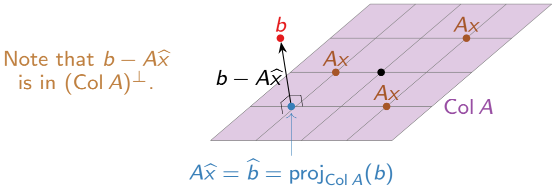
Least Squares Solutions

Definition

Let A be an $m \times n$ matrix. A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$A\widehat{x} = \widehat{b} = \text{proj}_{\text{Col } A}(b).$$

> A least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.



In *distance terms*, for all x in $\mathbf{R}^n$:

$$\|b - A\widehat{x}\| \leq \|b - Ax\|$$

Least Squares Solutions: General Solution

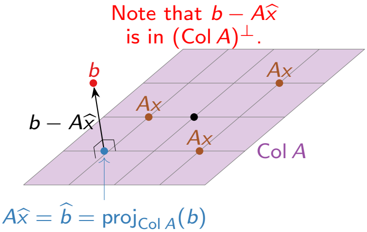
Theorem

Let A be a $m \times n$ **matrix**. Least squares solutions to $Ax = b$ are *any of the solutions to*

$$(A^T A)\widehat{x} = A^T b.$$

Now we can solve the problem without computing $\widehat{b}$ first.

> This is just another sysmtem of equations,
> but now it *is consistent* and uses *square matrix $A^T A$*!



Why is this true?

Recall: $(\text{Col}\,A)^\perp = \text{Nul}(A^T)$.

Now, $b - A\widehat{x}$ is in $(\text{Col}\,A)^\perp$ if and only if

$$A^T(b - A\widehat{x}) = 0.$$

In other words, $A^T A\widehat{x} = A^T b$.

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

First: Compute new matrix and vector

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

Second: Solve the new system; row reduce:

$$\left(\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -3 \end{array}\right).$$

So the *unique* least squares *solution is* $\widehat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

Find the least squares solutions to $Ax = b$ where:

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

First: Compute new matrix and vector

$$A^T A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

Second: Solve the new system; row reduce:

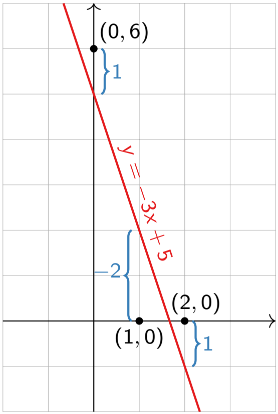
$$\left(\begin{array}{cc|c} 5 & -1 & 2 \\ -1 & 5 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array}\right).$$

So the *unique* least squares *solution is* $\widehat{x} = \begin{pmatrix} 1/3 \\ -1/3 \end{pmatrix}$.

Data Modeling: Best fit line

What does it minimize?

Best fit line minimizes the **sum of the squares** of the *vertical distances from the data points* to the line.



Data modeling: best fit parabola

What least squares problem $Ax = b$ finds **the best parabola** through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

The general equation for a parabola is

$$ax^2 + bx + c = y.$$

So we want to solve:

$$\begin{aligned} a(-1)^2 + b(-1) + c &= 0.5 \\ a(1)^2 + b(1) + c &= -1 \\ a(2)^2 + b(2) + c &= -0.5 \\ a(3)^2 + b(3) + c &= 2 \end{aligned}$$

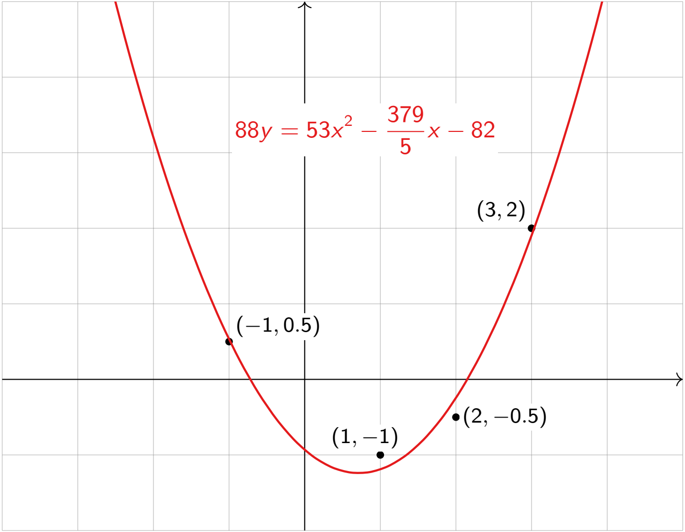
In matrix form:

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix}.$$

Answer: $\widehat{a} = \frac{53}{88}, \widehat{b} = \frac{379}{440}, \widehat{c} = \frac{82}{88}$ so best fit is: $53x^2 - \dfrac{379}{5}x - 82 = 88y$

Data modeling: best fit parabola

Picture

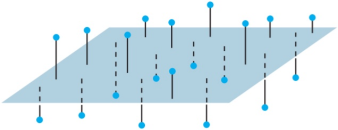


Multiple regression

Expert's notation

The model to fit:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 u_1 + \beta_2 v_1 + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 u_2 + \beta_2 v_2 + \epsilon_2 \\
&\vdots \qquad\qquad\qquad \vdots \\
y_n &= \beta_0 + \beta_1 u_n + \beta_2 v_n + \epsilon_n
\end{aligned}
$$



The equation display $y = X\beta + \varepsilon$:

Observation vector, Design matrix, Parameter vector, Residual vector

$$
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

The error

We want to *minimize the length of ε*.
In last section we don't write it as part of the equation.

Symmetric matrices

Definition

An $n \times n$ matrix is **symmetric** if $A = A^T$.

> **Theorem**
>
> An $n \times n$ matrix A is *orthogonally diagonalizable* if and only if A is *symmetric*.

The easy observation: Let $A = PDP^T$ with D diagonal and P orthonormal.

Just check A is symmetric, that is $A = A^T$:

$$(\underbrace{PDP^T}_{A}) = (P^T)^T D^T P^T = \underbrace{PDP^T}_{A}$$

The difficult part (omitted here) is: if $A = A^T$ then

an *orthogonal diagonalization do exists*.

Summary

Spectral Theorem for Symmetric matrices

An $n \times n$ *symmetric* matrix A has the **following properties**.

- A has n *real eigenvalues*, counting multiplicities
- For each eigenvalue,
  the dimension of the λ-eigenspaces equal the algebraic multiplicity.
- The eigenspaces are *mutually orthogonal!*
  eigenvectors corresponding to different eigenvalues are orthogonal.
- A is *orthogonally diagonalizable*.

Example: Orthogonally diagonalizable

Example

Orthogonally diagonalize the matrix $A = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$

its *charactheristic equation* is $-(\lambda - 7)^2(\lambda + 2) = 0$.

Find a basis for each λ-eigenspace:

For $\lambda = 7$: $\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix} \right\}$ For $\lambda = -2$: $\left\{ \begin{pmatrix} -1 \\ -1/2 \\ 1 \end{pmatrix} \right\}$

A suitable P

Is the set of eigenvectors above already orthogonal? orthonormal?

$$A = P \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix} P^{-1}$$

Example: Orthogonally diagonalizable
continued

Verify:

- $v_3 = \begin{pmatrix} -1 \\ -1/2 \\ 1 \end{pmatrix}$ is *already orthogonal* to $v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix}$
- but $v_1 \cdot v_2 \neq 0$.

Tackle this: Use Gram-Schmidt

$$u_1 = v_1$$

$$u_2 = v_2 - \frac{v_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix} - \frac{-1/2}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/4 \\ 1 \\ 1/4 \end{pmatrix}$$

And $u_3 = v_3$. **Then normalize!**

$$P = \begin{pmatrix} 1/\sqrt{2} & -1\sqrt{18} & -2/3 \\ 0 & 4\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1\sqrt{18} & 2/3 \end{pmatrix}, \qquad D = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Example: Spectral Decomposition

Example

Construct a *spectral decomposition* of the matrix A with orthogonal diagonalization

$$A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}$$

Solution: Then $A = 8u_1u_1^T + 3u_2u_2^T$, each matrix is

$$u_1u_1^T = \begin{pmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{pmatrix}$$

$$u_2u_2^T = \begin{pmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{pmatrix}$$

Check: $8u_1u_1^T + 3u_2u_2^T = \begin{pmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{pmatrix} + \begin{pmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{pmatrix} = A$

Back to change of variables

A consequence of the spectral theorem for symmetric matrices

The principal axes theorem

Let A be $n \times n$ symmetric matrix.

Then there is an **orthogonal change of variable** $x = Py$ that transforms the quadratic form $x^T Ax$ into a quadratic form

$y^T Dy$ with no cross-product terms.

If $A = PDP^{-1}$ with $P^T = P^{-1}$ and D diagonal, then

$$x^T Ax = \underbrace{x^T P}_{y^T} \; D \; \underbrace{P^{-1} x}_{y}$$

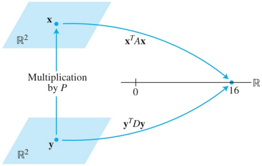


FIGURE 1 Change of variable in $x^T Ax$.

Change of variables

Example

Make a change of variables that transforms the quadratic form

$$Q(x_1, x_2) = x_1^2 - 5x_2^2 - 8x_1x_2$$

into a quadratic form with *no cross-product* terms

General Formula: there is an *orthonormal matrix P* such that

$$A = P \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} P^T$$

the change of variables is given by $y = P^T x = P^{-1} x$.

In this case, First $A = \begin{pmatrix} 1 & -4 \\ -4 & 5 \end{pmatrix}$, $\lambda_1 = 3, \lambda_2 = -7$ and $P = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$
Then

$$y^T \begin{pmatrix} 3 & 0 \\ 0 & -7 \end{pmatrix} y = 3y_1^2 - 7y_2^2$$

Classify quadratic forms

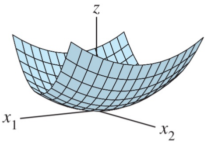
A quadratic form is

- *Indefinite:* if $Q(x)$ assumes **both** positive and negative values
- *Positive definite:* if $Q(x) > 0$ for all $x \neq 0$,
- *Negative definite:* if $Q(x) < 0$ for all $x \neq 0$,

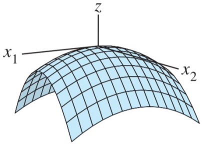
The prefix *semi* means e.g. $Q(x) \geq 0$ for all $x \neq 0$.

Eigenvalues

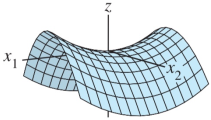
You can classify quadratic from knowing its eigenvalues (evaluate on principal axes)
e.g. Positive definite forms have *all eigenvalues* positive.



Positive definite



Negative definite



Indefinite