# Symmetric Matrices

## Geometric Algorithms
## Lecture 25

CAS CS 132

# Introduction

# Objectives

1. Finish up our discussion of linear models (actually define linear models).
2. Talk briefly about symmetric matrices and eigenvalues.
3. Describe an application to constrained optimization problems.

# Keywords

linear models

design matrices

general linear regression

symmetric matrices

the spectral theorem

orthogonal diagonalizability

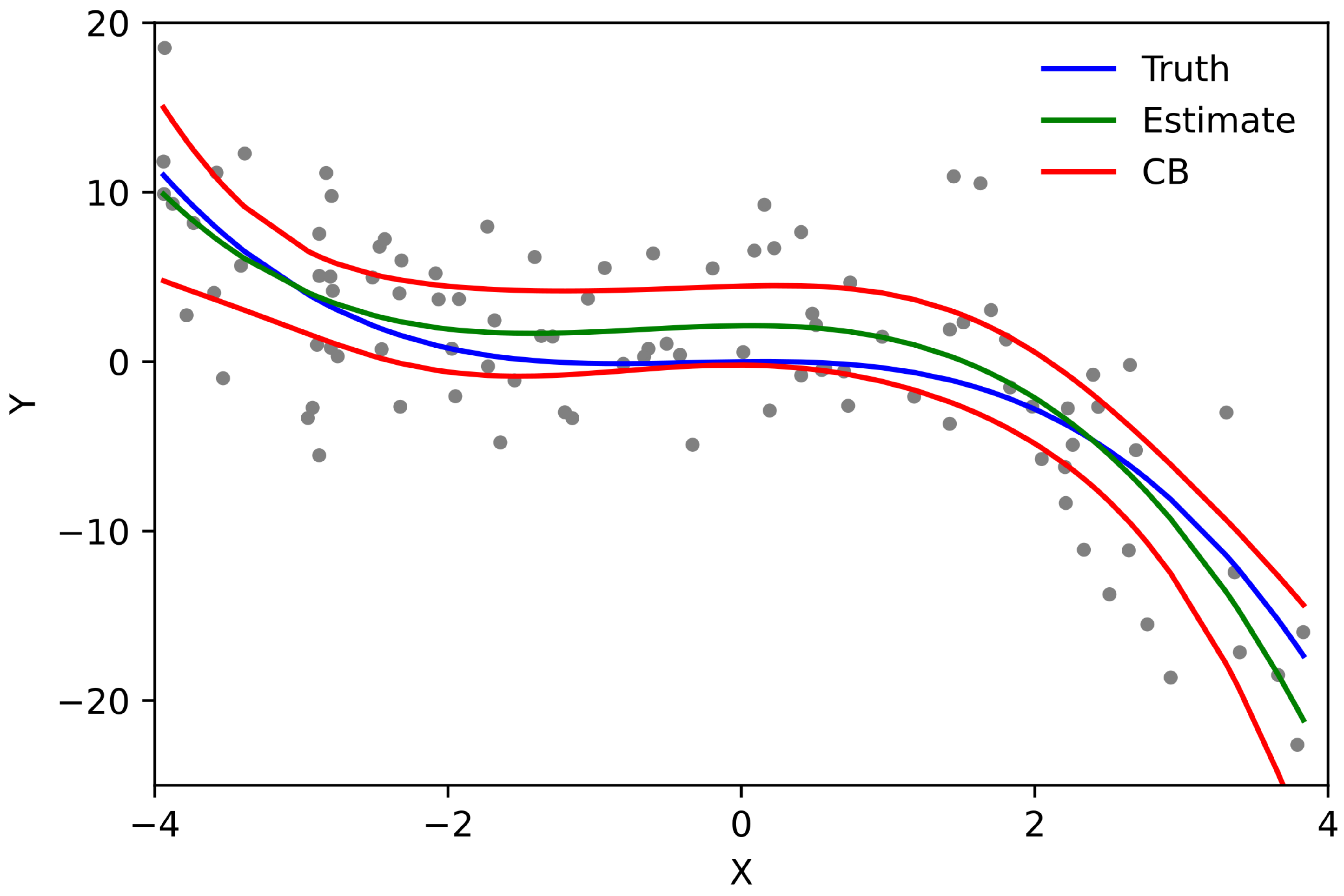
quadratic forms

definiteness

constrained optimization

# Recap

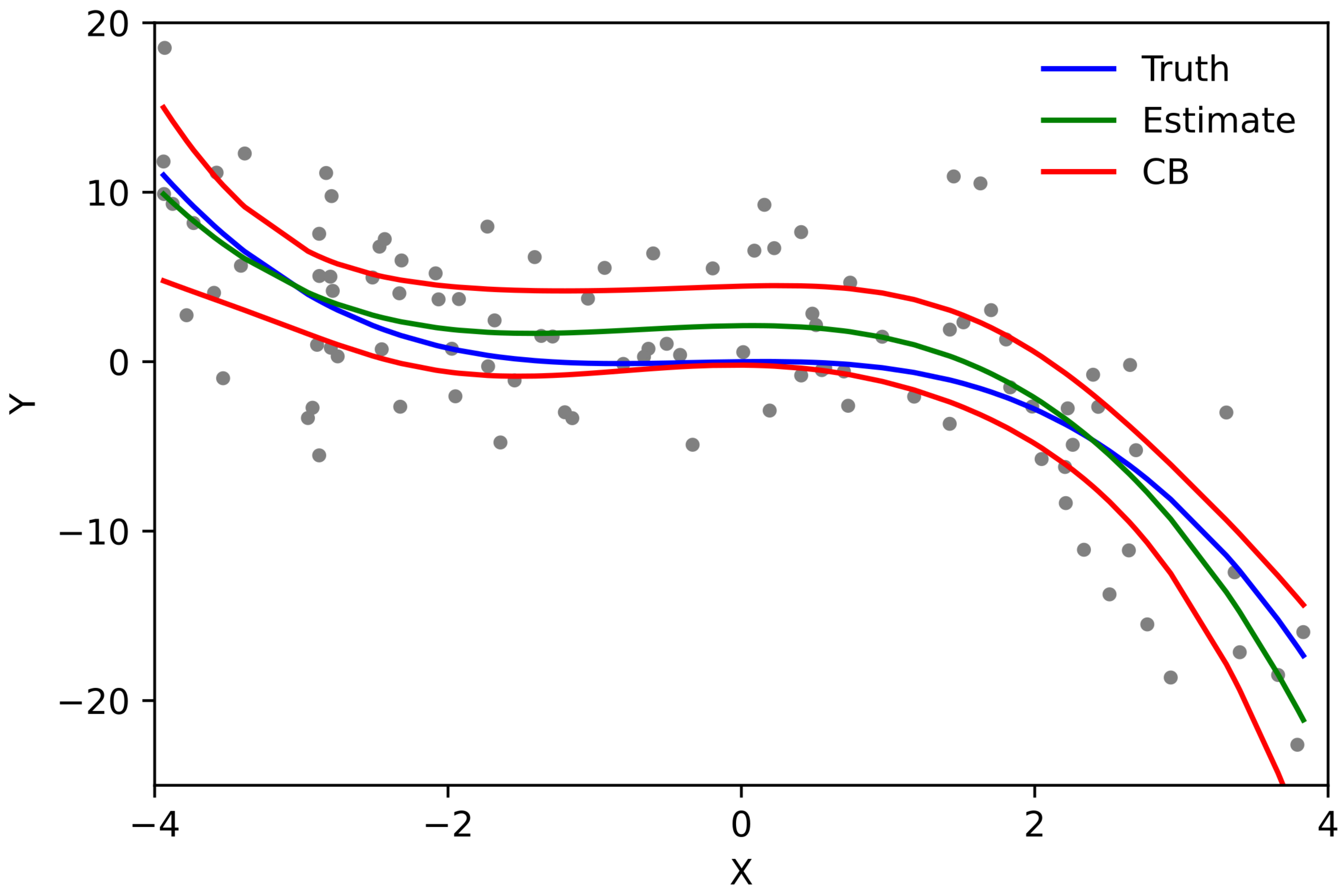
# Recall: General Regression




https://commons.wikimedia.org/wiki/File:Polyreg_scheffe.svg

# Recall: General Regression

**Regression** is the process of estimating the relationships independent and dependent variables in a dataset.

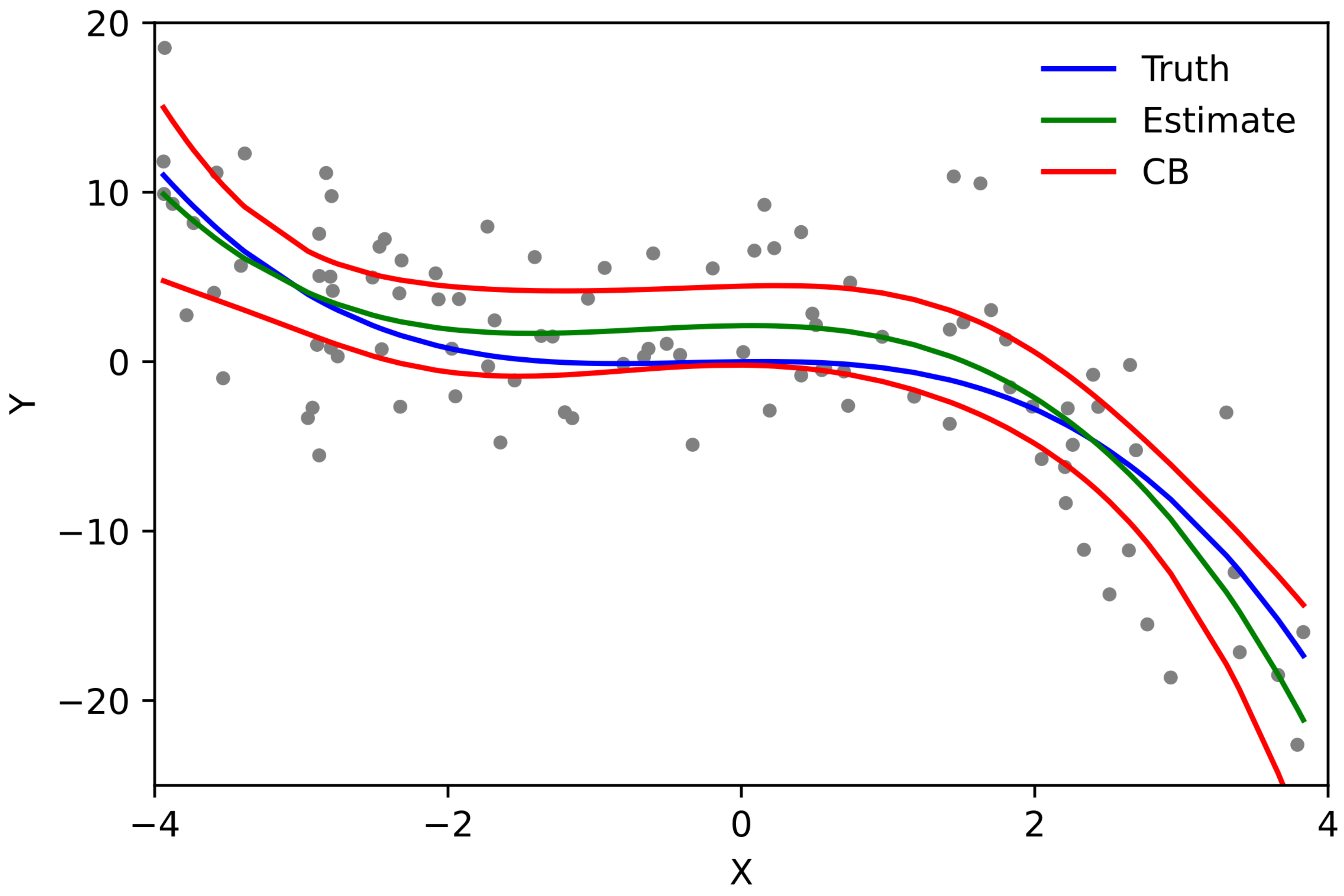



https://commons.wikimedia.org/wiki/File:Polyreg_scheffe.svg

# Recall: General Regression

**Regression** is the process of estimating the relationships independent and dependent variables in a dataset.

What we are estimating is a <u>mathematical function</u>



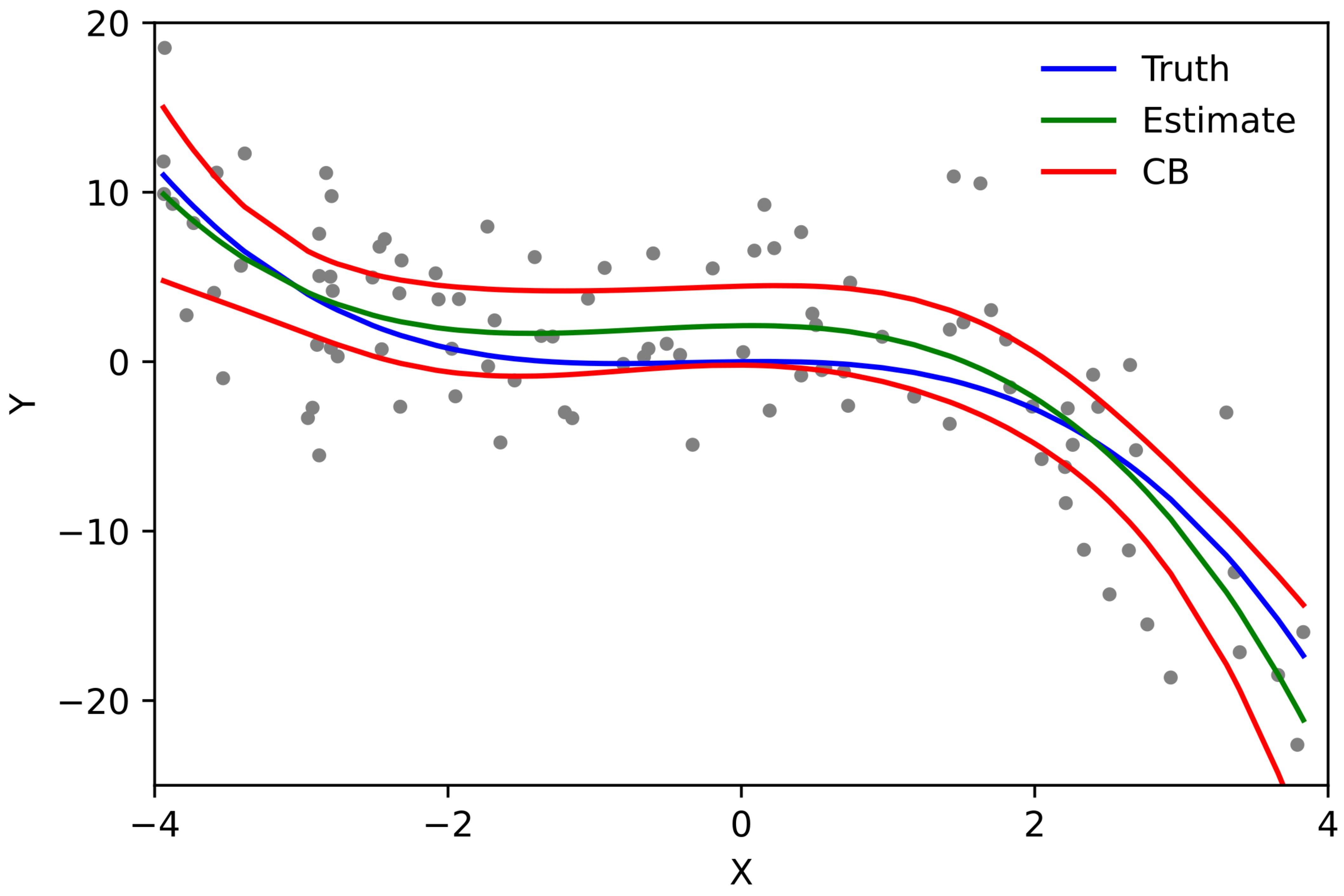

https://commons.wikimedia.org/wiki/File:Polyreg_scheffe.svg

# Recall: General Regression

**Regression** is the process of estimating the relationships independent and dependent variables in a dataset.

What we are estimating is a <u>mathematical function</u>

We think of the environment has providing us a function from our independent variables to our dependent variables.




https://commons.wikimedia.org/wiki/File:Polyreg_scheffe.svg

# Recall: How To: Line of Best Fit

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

# Recall: How To: Line of Best Fit

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

**Problem.** Find the least squares line for the dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

# Recall: How To: Line of Best Fit

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$y = \beta_0 + \beta_1 x$

**Problem.** Find the least squares line for the dataset $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

**Solution. Find the least squares solution to the above equation.**

# Recall: "Vectors" of Generalization

# Recall: "Vectors" of Generalization

1. What if we have *more than one* independent value?

# Recall: "Vectors" of Generalization

1. What if we have *more than one* independent value?

**multiple regression, (hyper)plane of best fit**

# Recall: "Vectors" of Generalization

1. What if we have *more than one* independent value?

   **multiple regression, (hyper)plane of best fit**

2. What if our data is not *exactly* linear.

# Recall: "Vectors" of Generalization

1. What if we have *more than one* independent value?

   **multiple regression, (hyper)plane of best fit**

2. What if our data is not *exactly* linear.

   **e.g., polynomial regression**

# Recall: Plane of Best Fit

**Dataset:** $\{(x_1, y_1, z_1), \ldots, (x_k, y_k, z_k)\}$ where $(x_i, y_i)$ is an longitude and latitude and $z_i$ is an altitude.

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x, y) = \beta_0 + \beta_1 x + \beta_2 y$$

which minimizes

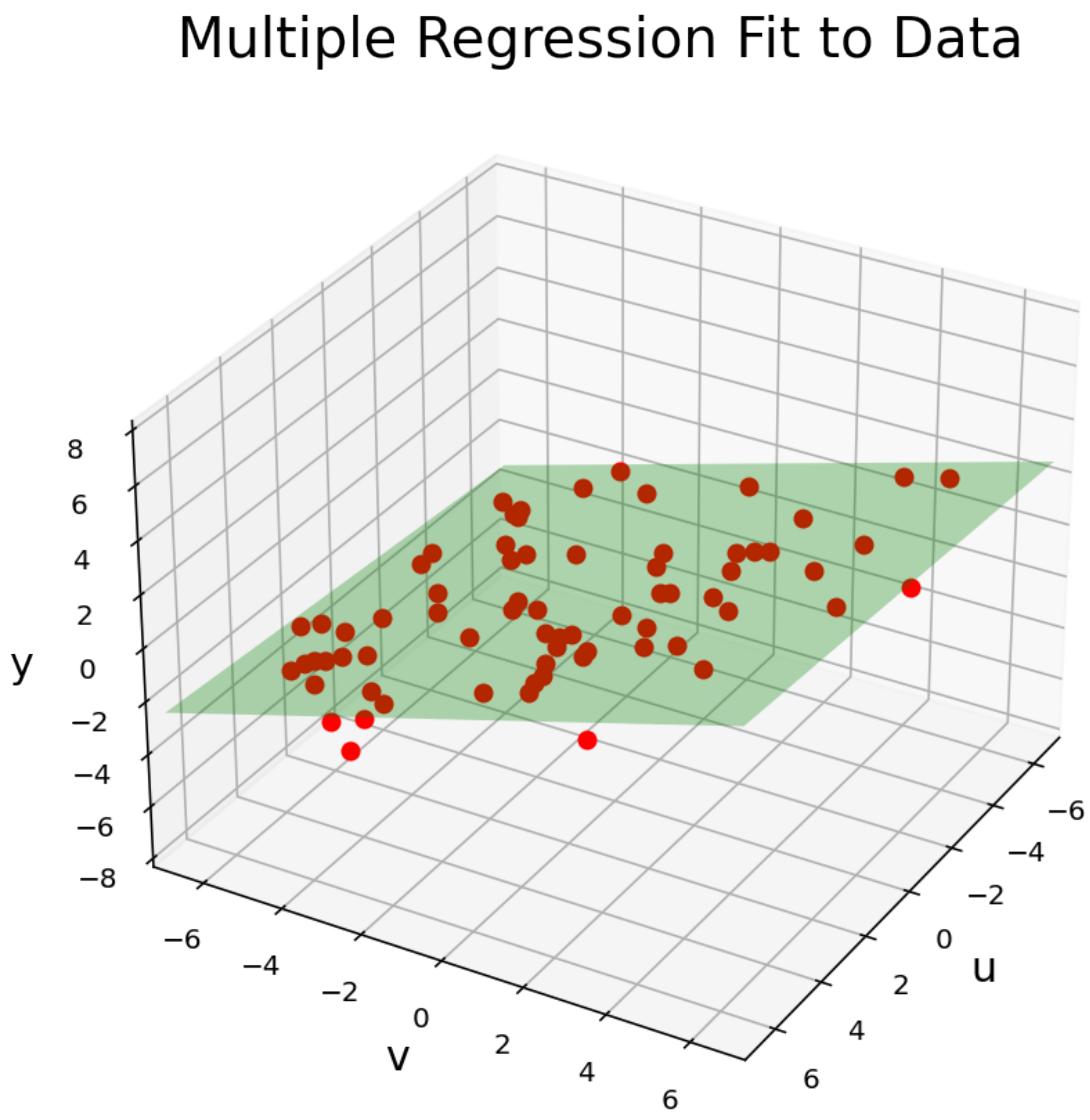
$$\sum_{i=1}^{k} (f(x_i, y_i) - z_i)^2$$

residual

prediction

observation

Figure 23.2



# Recall: Plane of Best Fit

**Dataset:** $\{(x_1, y_1, z_1), \ldots, (x_k, y_k, z_k)\}$
where $(x_i, y_i)$ is an longitude and latitude and $z_i$ is an altitude.

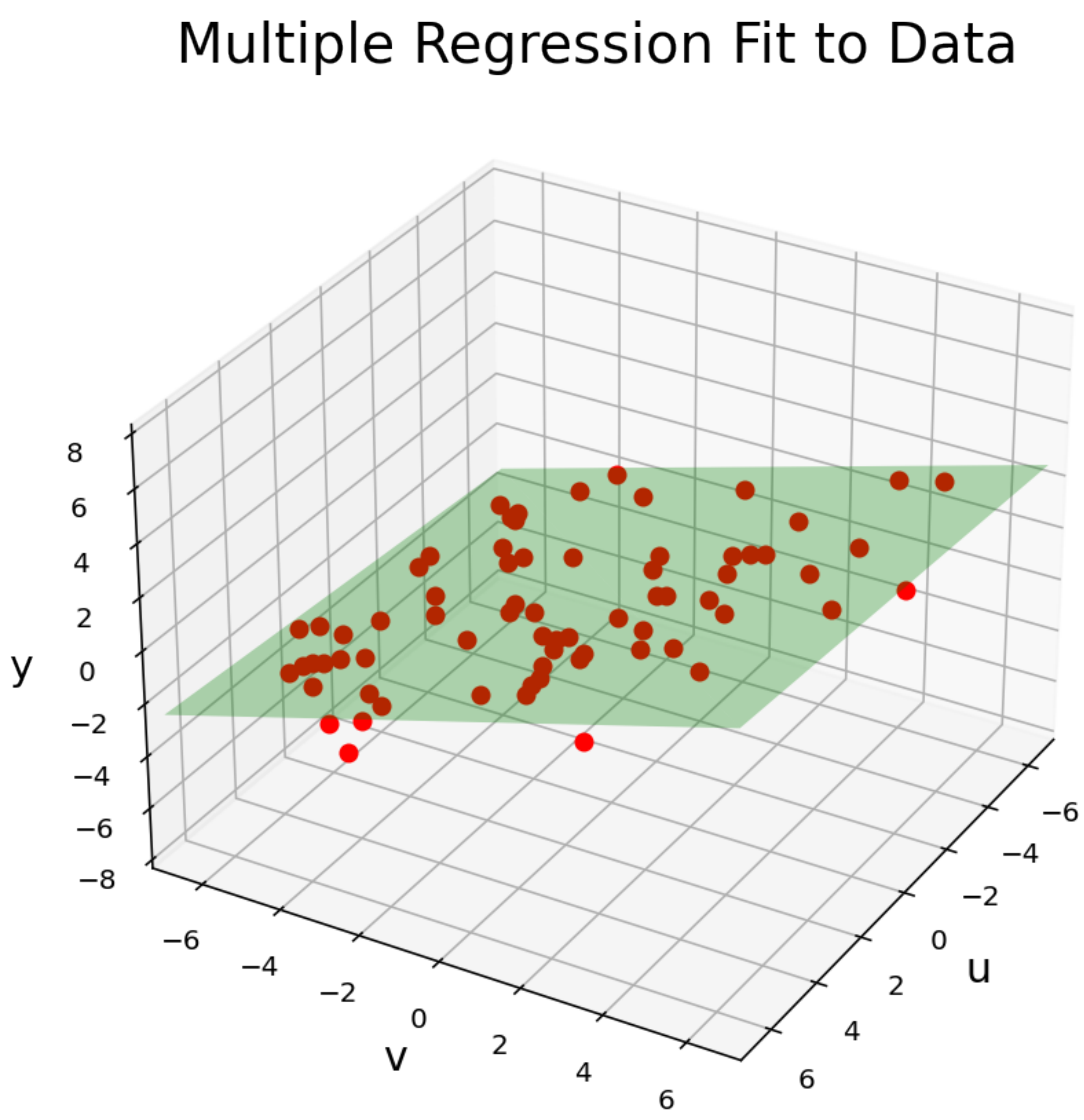
**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x, y) = \beta_0 + \beta_1 x + \beta_2 y$$

which minimizes

$$\sum_{i=1}^{k} (f(x_i, y_i) - z_i)^2$$

Figure 23.2




***$f(x, y)$ is a good approximation of the altitude.***

# Recall: Plane of Best Fit

**Dataset:** $\{(x_1, y_1, z_1), \ldots, (x_k, y_k, z_k)\}$
where $(x_i, y_i)$ is an longitude and latitude and $z_i$ is an altitude.

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

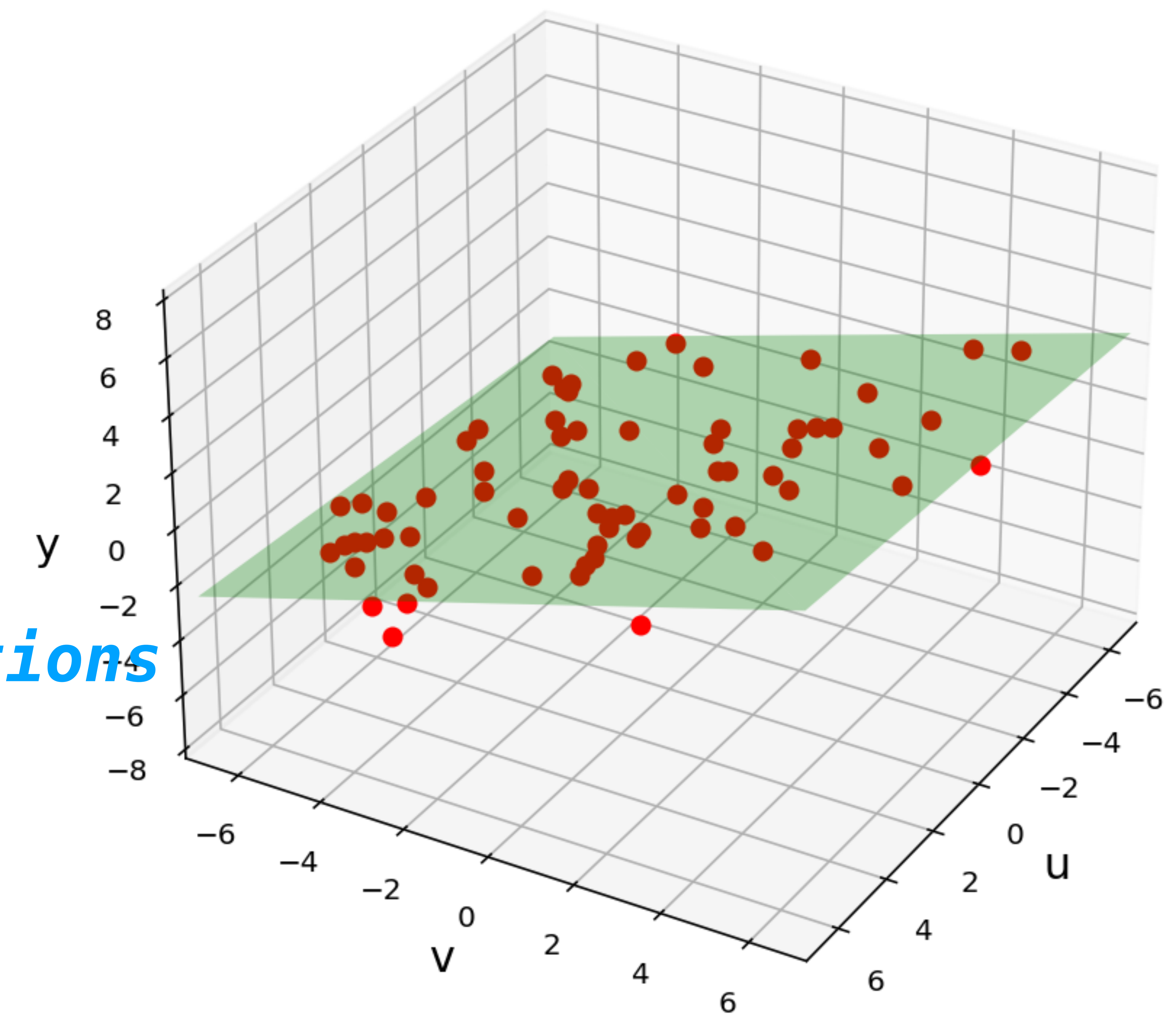
$$f(x, y) = \beta_0 + \beta_1 x + \beta_2 y$$

*recall: planes are given by linear equations*

which minimizes

$$\sum_{i=1}^{k} (f(x_i, y_i) - z_i)^2$$

Figure 23.2




*$f(x, y)$ is a good approximation of the altitude.*

# Recall: Parabola of Best Fit

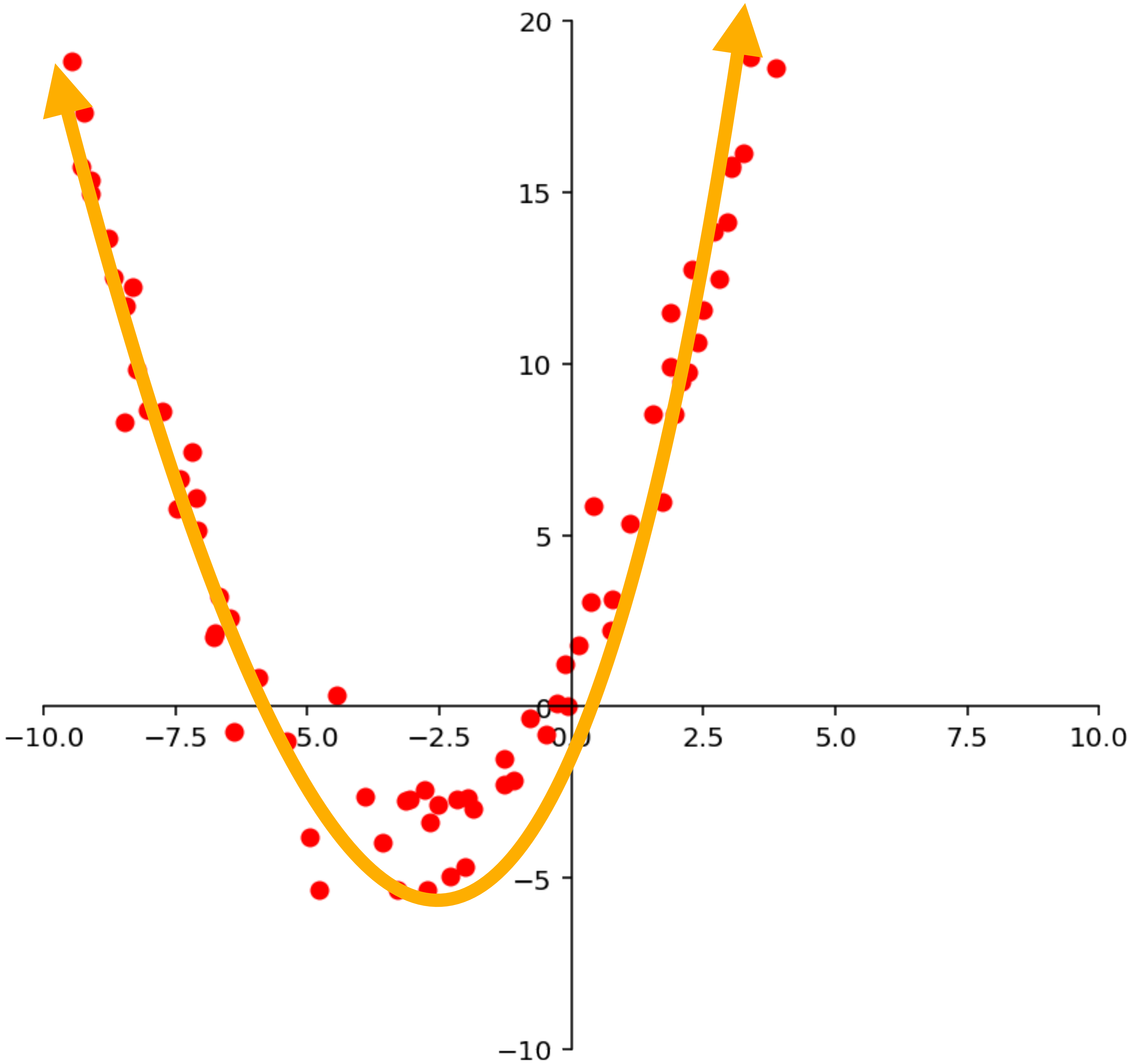
**Dataset:** $\{(x_1, y_1), \ldots, (x_k, y_k)\}$

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

minimizes

$$\sum_{i=1}^{k} (f(x_i) - y_i)^2$$



# Recall: Parabola of Best Fit

**Dataset:** $\{(x_1, y_1), \ldots, (x_k, y_k)\}$

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

minimizes

$$\sum_{i=1}^{k} (f(x_i) - y_i)^2$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 = y_1$$
$$\beta_0 + \beta_1 x_2 + \beta_2 x_2^2 = y_2$$
$$\vdots$$
$$\beta_0 + \beta_1 x_k + \beta_2 x_k^2 = y_k$$

**Step 1:** Set up an (almost assuredly inconsistent) system of linear equations in terms of the variables $\beta_0, \beta_1, \beta_2$

# Recall: Parabola of Best Fit

This is still linear in the $\beta$'s

**Dataset:** $\{(x_1, y_1), \ldots, (x_k, y_k)\}$

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

minimizes

$$\sum_{i=1}^{k} (f(x_i) - y_i)^2$$

linear equation

$$\beta_0 + \beta_1 x_1 + \beta_2 x_1^2 = y_1$$
$$\beta_0 + \beta_1 x_2 + \beta_2 x_2^2 = y_2$$
$$\vdots$$
$$\beta_0 + \beta_1 x_k + \beta_2 x_k^2 = y_k$$

**Step 1:** Set up an (almost assuredly inconsistent) system of linear equations in terms of the variables $\beta_0, \beta_1, \beta_2$

# Recall: Parabola of Best Fit

**Dataset:** $\{(x_1, y_1), \ldots, (x_k, y_k)\}$

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

minimizes

$$\sum_{i=1}^{k} (f(x_i) - y_i)^2$$

$$\overset{X}{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_k & x_k^2 \end{bmatrix}} \overset{\vec{\beta}}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}} = \overset{\mathbf{y}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}}$$

**Step 2:** Rewrite the system as a matrix equation.

# Recall: Parabola of Best Fit

**Dataset:** $\{(x_1, y_1), \ldots, (x_k, y_k)\}$

**Problem:** Find $\beta_0, \beta_1, \beta_2$ such that

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

minimizes

$$\sum_{i=1}^{k} (f(x_i) - y_i)^2$$

error

least squares solution

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$

**Step 3:** Find the least squares solution of this system and use as the parameters of your model.

# Recap Problem

$$\{(0,3),(1,1),(-1,1),(2,3)\}$$

*Find the matrices $X$ as in the previous example to find the least squares best fix parabola <u>and the least squares best fit cubic</u> for this dataset.*

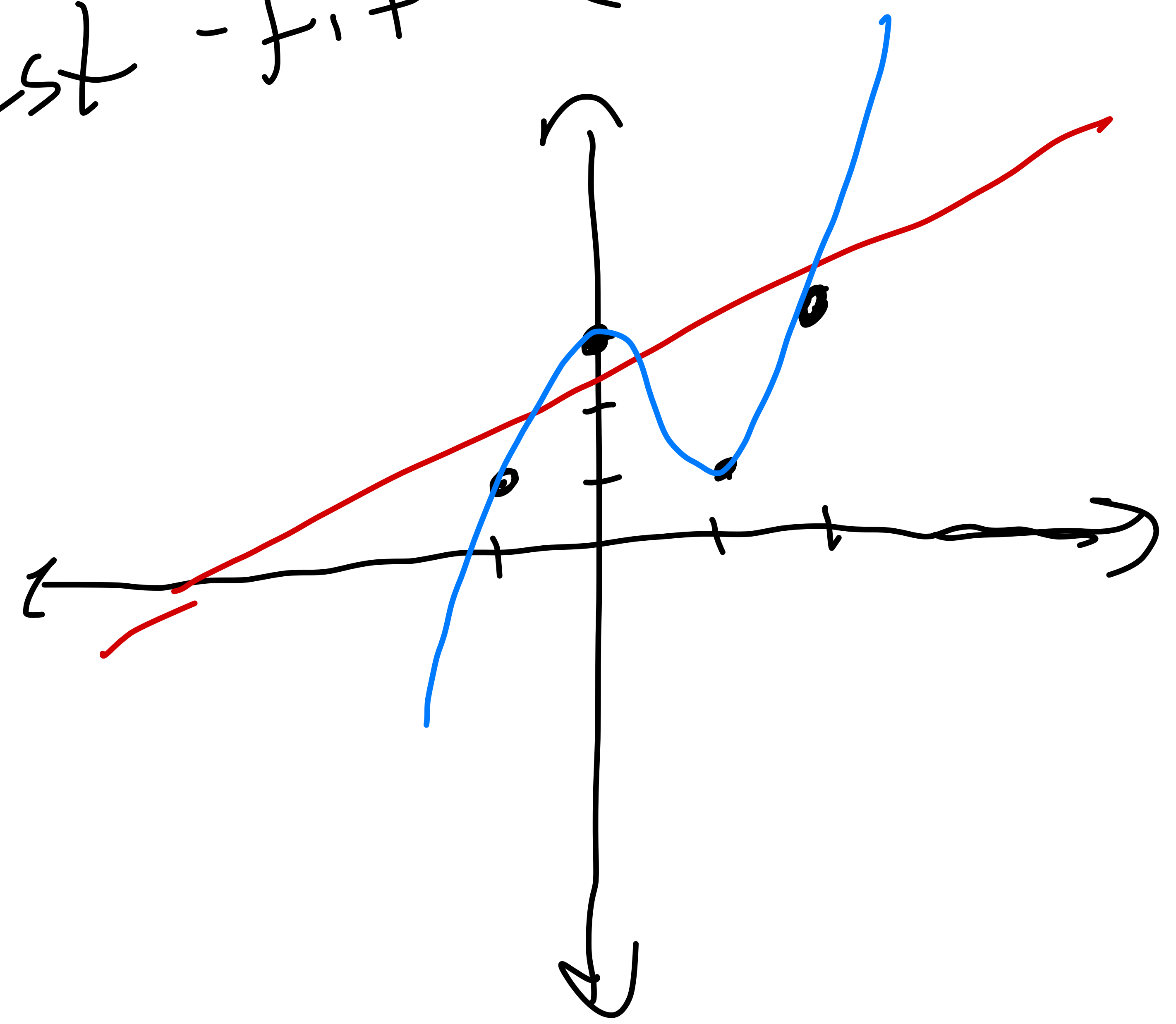
# Answer

$\{(0,3), (1,1), (-1,1), (2,3)\}$

$$\begin{bmatrix} & \vdots & \\ 1 & x_i & x_i^2 \\ & \vdots & \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$ for best-fit parabola

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

$$\begin{bmatrix} & \vdots & & \\ 1 & x_i & x_i^2 & x_i^3 \\ & \vdots & & \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{bmatrix}$$ for best-fit cubic



# Design Matrices

# The Takeaway

We can use non-linear modeling functions as long as they are linear in the parameters.

# Linear in Parameters

vectors     number

**Definition.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is **linear in the parameters** $\beta_1, \ldots, \beta_k$ if it can be written as

$$f(\mathbf{x}) = \beta_1 \phi_1(\mathbf{x}) + \beta_2 \phi_2(\mathbf{x}) + \ldots + \beta_k \phi_k(\mathbf{x})$$

not necessarily linear

for functions $\phi_1, \ldots, \phi_k : \mathbb{R}^n \to \mathbb{R}$

Example: $f(x_1, x_2) = \beta_0 e^{x_1} + \beta_1 \ln(x_1 \cdot x_2)$

Nonexamples:

$f(x_1, x_2) = \beta_0 \beta_1 x_1 x_2$     $f(x_1) = e^{\beta_0 x_1}$

# An Aside: Statistical Models (Another view)

$$\mathbf{y} = X\vec{\beta} + \vec{\epsilon}$$

# An Aside: Statistical Models (Another view)

$$\mathbf{y} = X\vec{\beta} + \vec{\epsilon}$$

So far, we have been considering *inconsistent* systems of the form $\mathbf{y} = X\vec{\beta}$.

# An Aside: Statistical Models (Another view)

$$\mathbf{y} = X\vec{\beta} + \vec{\epsilon}$$

So far, we have been considering *inconsistent* systems of the form $\mathbf{y} = X\vec{\beta}$.

It is also common to *make the system consistent* by adding error terms (the $\epsilon$'s).

# An Aside: Statistical Models (Another view)

$$\mathbf{y} = X\vec{\beta} + \vec{\epsilon}$$

So far, we have been considering *inconsistent* systems of the form $\mathbf{y} = X\vec{\beta}$.

It is also common to *make the system consistent* by adding error terms (the $\epsilon$'s).

*(We won't use this view, this is mostly for your personal betterment, and because the notes use this notation occasionally.)*

# An Aside: Statistical Models (Another view)

design matrix

$$\mathbf{y} = X\vec{\beta} + \vec{\epsilon}$$

So far, we have been considering *inconsistent* systems of the form $\mathbf{y} = X\vec{\beta}$.

It is also common to *make the system consistent* by adding error terms (the $\epsilon$'s).

*(We won't use this view, this is mostly for your personal betterment, and because the notes use this notation occasionally.)*

# The Takeaway (Again)

We can build design matrices for function which are linear in their parameters.

# Linear (Regression) Model

**Definition.** A **linear model** with parameters $\beta_1, \ldots, \beta_k$ is a function $f : \mathbb{R}^n \to \mathbb{R}$ which is linear in the parameters $\beta_1, \ldots, \beta_k$.

The *model fitting problem* is the problem of determining which parameters fit the data "best".

# General Linear Regression

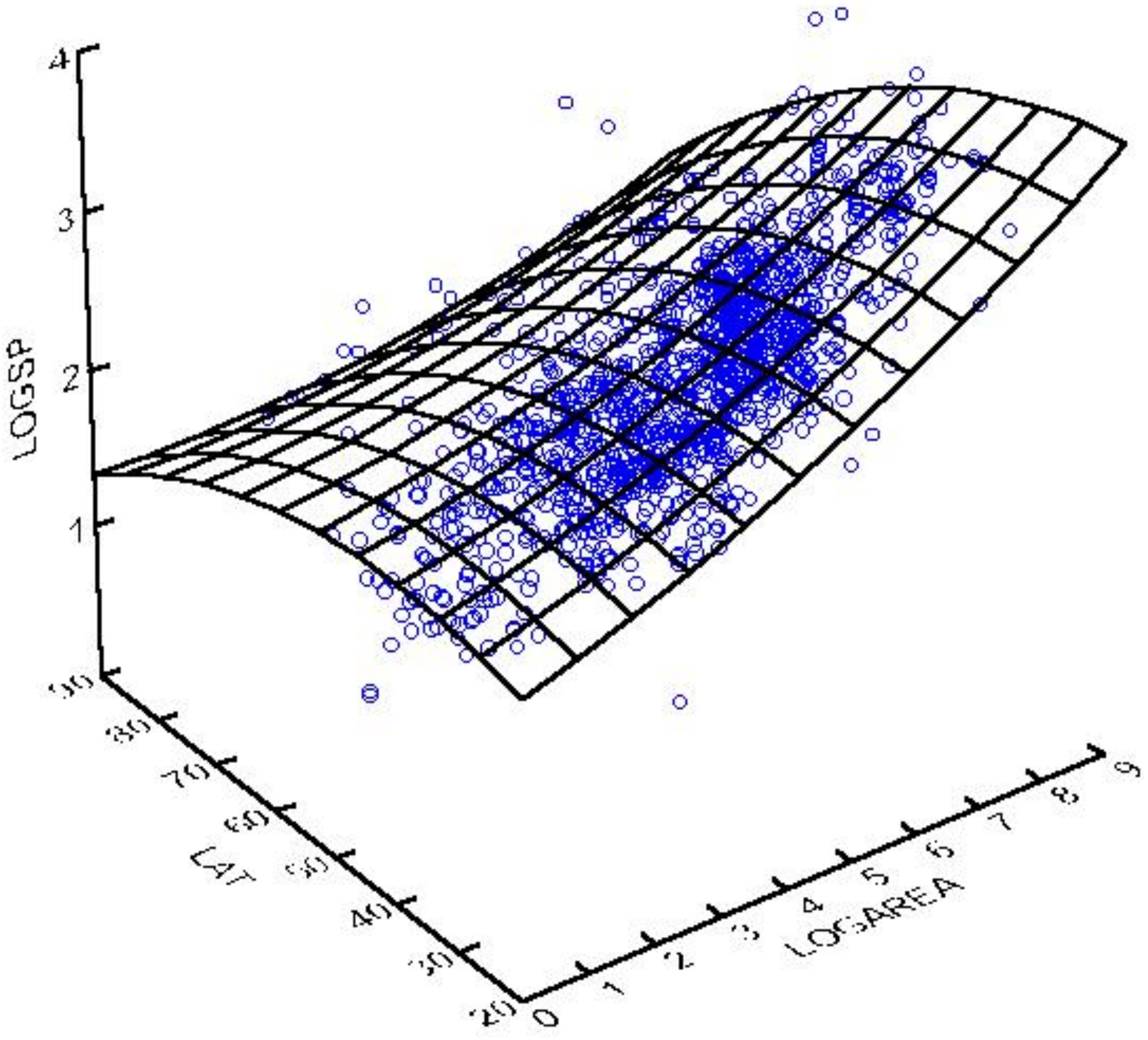
**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$




https://ordination.okstate.edu/MULTIPLE.htm

# General Linear Regression

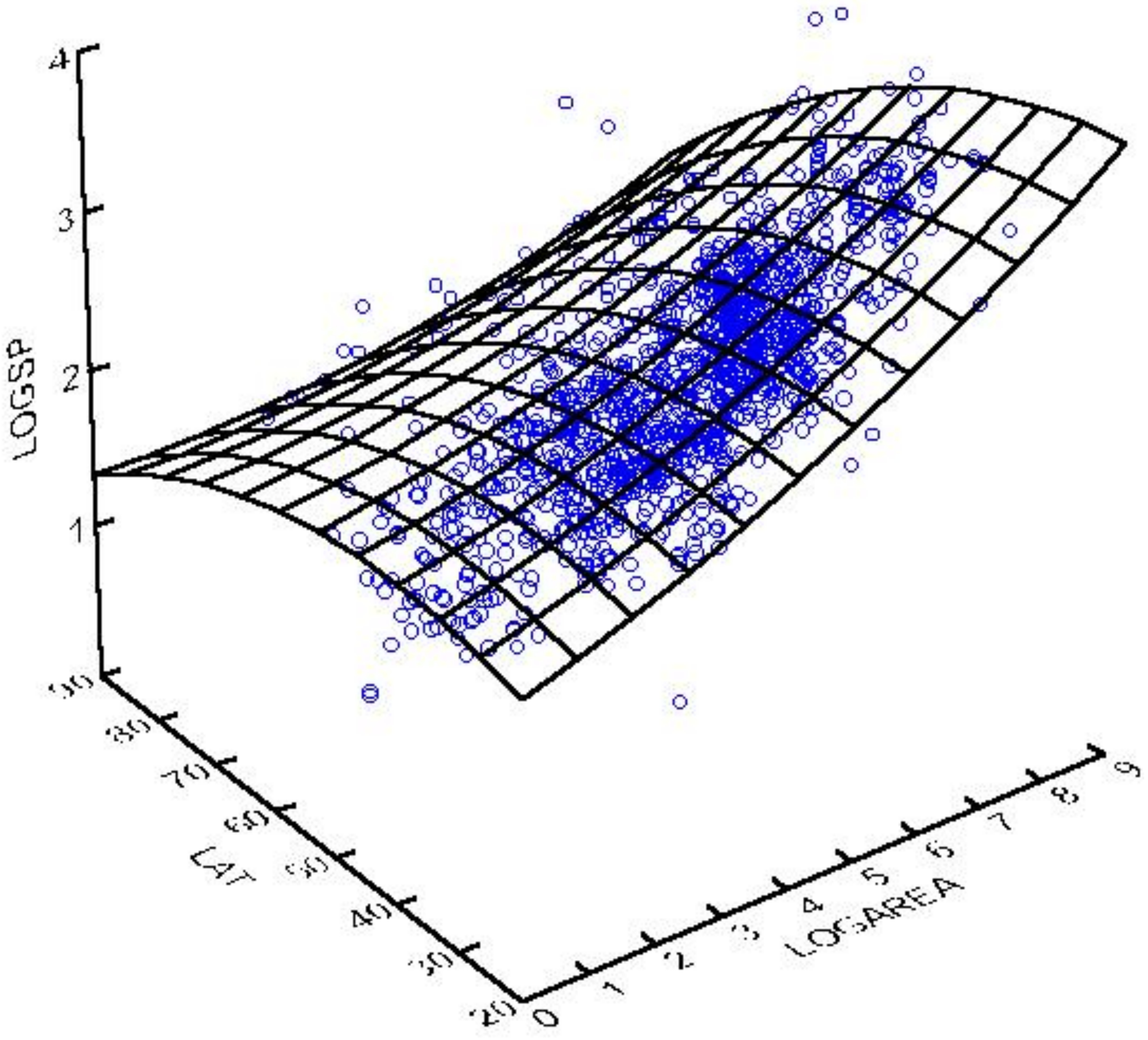
**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$




**Build a linear model which minimizes the least-squares error.**

https://ordination.okstate.edu/MULTIPLE.htm

# General Linear Regression

**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$

$$\beta_1\phi_1(\mathbf{x}_1) + \ldots + \beta_k\phi_k(\mathbf{x}_1) = y_1$$
$$\beta_1\phi_1(\mathbf{x}_2) + \ldots + \beta_k\phi_k(\mathbf{x}_2) = y_2$$
$$\vdots$$
$$\beta_1\phi_1(\mathbf{x}_m) + \ldots + \beta_k\phi_k(\mathbf{x}_m) = y_2$$

**Step 1:** Set up an (almost assuredly inconsistent) system of linear equations in terms of the variables $\beta_1, \ldots, \beta_k$

# General Linear Regression

linear equation

**This is still linear in the $\beta$'s**

**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$

$$\beta_1\phi_1(\mathbf{x}_1) + \ldots + \beta_k\phi_k(\mathbf{x}_1) = y_1$$

$$\beta_1\phi_1(\mathbf{x}_2) + \ldots + \beta_k\phi_k(\mathbf{x}_2) = y_2$$

$$\vdots$$

$$\beta_1\phi_1(\mathbf{x}_2) + \ldots + \beta_k\phi_k(\mathbf{x}_2) = y_2$$

**Step 1:** Set up an (almost assuredly inconsistent) system of linear equations in terms of the variables $\beta_1, \ldots, \beta_k$

# General Linear Regression

**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$

design matrix

$$\overset{X}{\begin{bmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \ldots & \phi_k(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \ldots & \phi_k(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_m) & \phi_2(\mathbf{x}_m) & \ldots & \phi_k(\mathbf{x}_m) \end{bmatrix}} \overset{\vec{\beta}}{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \; \beta_k \end{bmatrix}} = \overset{\mathbf{y}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{bmatrix}}$$

**Step 2:** Rewrite the system as a matrix equation.

# General Linear Regression

**dataset:** $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$

**Problem.** Given a function

$$f_{\beta_1,\ldots,\beta_k} : \mathbb{R}^n \to \mathbb{R}$$

which is *linear in the parameters* $\beta_1, \ldots \beta_k$, find values for $\beta_1, \ldots, \beta_k$ which minimize

$$\sum_{i=1}^{k} (f_{\beta_1,\ldots,\beta_k}(\mathbf{x}_i) - y_i)^2$$

$$\hat{\vec{\beta}} = (X^T X)^{-1} X^T \mathbf{y}$$

**Step 3:** Find the least squares solution of this system and use as the parameters of your model.

# How To: Design Matrices

# How To: Design Matrices

**Problem.** Find the design matrix for least squares regression with the function $f$ in terms of the parameters $\beta_1, \beta_2, \ldots, \beta_k$ given the dataset $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$.

# How To: Design Matrices

**Problem.** Find the design matrix for least squares regression with the function $f$ in terms of the parameters $\beta_1, \beta_2, \ldots, \beta_k$ given the dataset $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$.

**Solution.** First write $f(\mathbf{x})$ as $\beta_1 \phi_1(\mathbf{x}) + \ldots + \beta_k \phi(\mathbf{x})$ where $\phi_1, \ldots, \phi_k$ are potentially non-linear functions. Then build the matrix:

$$\begin{bmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \ldots & \phi_k(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \ldots & \phi_k(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_m) & \phi_2(\mathbf{x}_m) & \ldots & \phi_k(\mathbf{x}_m) \end{bmatrix}$$

# Question

*Find the design matrix for the least squares regression with the function*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \beta_1 \cos(x_1) + \beta_2 e^{-x_1 x_2} - \beta_1 x_3 + \beta_3$$

$$\beta_1 \left( \cos(x_1) - x_3 \right) + \dots$$

*for the dataset*

$$\mathbf{x}_1 = (0,0,0) \qquad y_1 = 5$$

$$\mathbf{x}_2 = (\pi,3,1) \qquad y_2 = 3$$

**Answer:** $\begin{bmatrix} 1 & 1 & 1 \\ -2 & e^{-3\pi} & 1 \end{bmatrix}$

$(0, 0, 0)$

$(\pi, 3, 1)$

$$f(x_1, x_2, x_3) = \beta_1 (\cos(x_1) - x_3) + \beta_2 e^{-x_1 x_2} + \beta_3 \, 1$$

$$\begin{bmatrix} \cos(x_1) - x_3 & e^{-x_1 x_2} & 1 \end{bmatrix}$$

$$\begin{bmatrix} \cos(0) - 0 & e^{0} & 1 \\ \cos(\pi) - 1 & e^{-3\pi} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 \\ -2 & e^{-3\pi} & 1 \end{bmatrix}$$

# Practical Considerations

# Practical Considerations

Many functions require large design matrices, e.g. multivariate polynomials have *a lot* of possible terms.

# Practical Considerations

Many functions require large design matrices, e.g. multivariate polynomials have *a lot* of possible terms.

We haven't actually talked about *which* modeling functions to use.

# Practical Considerations

Many functions require large design matrices, e.g. multivariate polynomials have *a lot* of possible terms.

We haven't actually talked about *which* modeling functions to use.

Again, is least–squares error really what we want? What if we want to minimize something else?

# Practical Considerations

Many functions require large design matrices, e.g. multivariate polynomials have *a lot* of possible terms.

We haven't actually talked about *which* modeling functions to use.

Again, is least-squares error really what we want? What if we want to minimize something else?

**Concerns for another class.**

# One Last Thing

Read though the extended example in the notes on "Multiple Regression in Practice."

It should be useful for Homework 12.

# Symmetric Matrices

# Recall: Symmetric Matrices

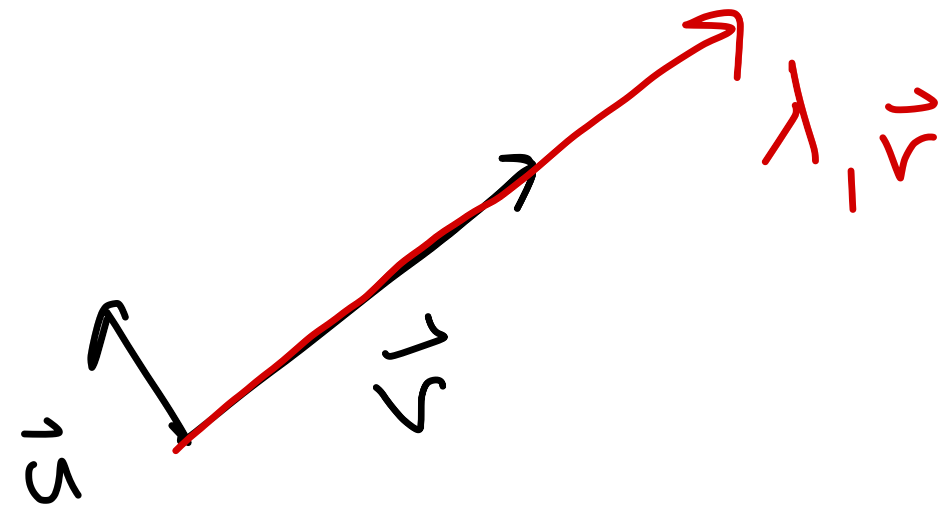
**Definition.** A square matrix $A$ is **symmetric** if $A^T = A$.

Example:

$$A_{ij} = A_{ji}$$

$$\begin{bmatrix} 5 & 0 & -2 \\ 0 & 2 & 5 \\ -2 & 5 & 3 \end{bmatrix}$$

# Orthogonal Eigenvectors




**Theorem.** For a symmetric matrix $A$, if $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors for *distinct* eigenvalues, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Verify: $A\vec u = \lambda_1 \vec u \qquad A\vec v = \lambda_2 \vec v$

WTS $\langle \vec u, \vec v \rangle = \vec 0$

$$\langle \vec u, A\vec v\rangle = u^T A v = u^T \lambda_2 \vec v = \lambda_2 \langle \vec u, \vec v\rangle$$

$$= u^T A^T \vec v = \langle A\vec u, \vec v\rangle = \lambda_1 \langle \vec u, \vec v\rangle$$

$\lambda_1 \neq \lambda_2$

$$(\lambda_1 - \lambda_2)\langle \vec u, \vec v\rangle = 0$$

$\lambda_1 - \lambda_2 \neq 0$, $\langle \vec u, \vec v\rangle = 0$

# Recall: Diagonalizable Matrices

# Recall: Diagonalizable Matrices

**Definition.** A matrix $A$ is **diagonalizable** if it is similar to a diagonal matrix.

# Recall: Diagonalizable Matrices

**Definition.** A matrix $A$ is **diagonalizable** if it is similar to a diagonal matrix.

*There is an invertible matrix $P$ and <u>diagonal</u> matrix $D$ such that $A = PDP^{-1}$.*

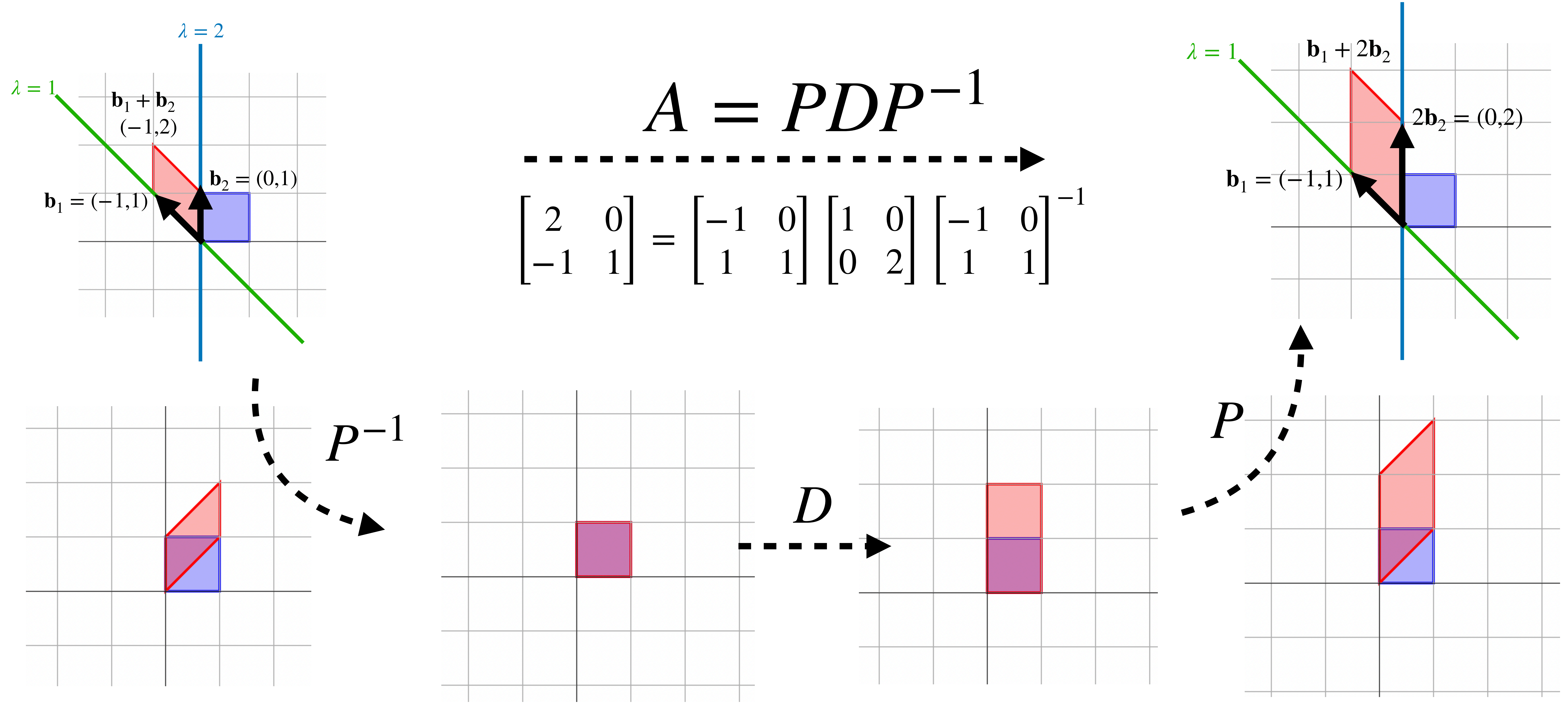
# Recall: Diagonalizable Matrices

**Definition.** A matrix $A$ is **diagonalizable** if it is similar to a diagonal matrix.

*There is an invertible matrix $P$ and diagonal matrix $D$ such that $A = PDP^{-1}$.*

**Diagonalizable matrices are the same as scaling matrices up to a change of basis.**

# Recall: The Picture



# Recall: The Diagonalization Theorem

$$A = PDP^{-1}$$

# Recall: The Diagonalization Theorem

$$A = PDP^{-1}$$

**Theorem.** $A$ is diagonalizable if and only if it has an eigenbasis.

# Recall: The Diagonalization Theorem

$$A = PDP^{-1}$$

**Theorem.** $A$ is diagonalizable if and only if it has an eigenbasis.

**The idea:**

# Recall: The Diagonalization Theorem

eigenbasis

$$A = PDP^{-1}$$

**Theorem.** $A$ is diagonalizable if and only if it has an eigenbasis.

**The idea:**

The columns of $P$ form an eigenbasis for $A$.

# Recall: The Diagonalization Theorem

eigenbasis

$$A = PDP^{-1}$$

eigenvalues

**Theorem.** $A$ is diagonalizable if and only if it has an eigenbasis.

**The idea:**

The columns of $P$ form an eigenbasis for $A$.

The diagonal of $D$ are the eigenvalues for each column of $P$.

# Recall: The Diagonalization Theorem

eigenbasis

$$A = PDP^{-1}$$

eigenvalues

**Theorem.** $A$ is diagonalizable if and only if it has an eigenbasis.

**The idea:**

The columns of $P$ form an eigenbasis for $A$.

The diagonal of $D$ are the eigenvalues for each column of $P$.

**The matrix $P^{-1}$ is a change of basis to this eigenbasis of $A$.**

# The Spectral Theorem

**Theorem.** If $A$ is symmetric, then it has an ***orthonormal*** eigenbasis.

(we won't prove this)

Symmetric matrices are diagonalizable.

But more than that, we can choose $P$ to be ***orthogonal***.

# Recall: Orthonormal Matrices

**Definition.** A matrix is **orthonormal** if its columns form an orthonormal set.

The notes call a square orthonormal matrix an **orthogonal** matrix.

# Recall: Inverses of Orthogonal Matrices

**Theorem.** If an $n \times n$ matrix $U$ is orthogonal (square orthonormal) then it is invertible and

$$U^{-1} = U^T$$

Verify:

# Orthogonal Diagonalizability

**Definition.** A matrix $A$ is **orthogonally diagonalizable** if there is a diagonal matrix $D$ and matrix $P$ such that

$$A = PDP^T = PDP^{-1}$$

$P$ must be an orthogonal matrix.

**Symmetric matrices are orthogonally diagonalizable**

# Orthogonal Diagonalizability and Symmetry

**Fact.** All orthogonally diagonalizable matrices are symmetric.

Verify:

$$A = PDP^T$$

$$A^T = ((PD)P^T)^T$$

$$= P^{TT}(PD)^T$$

$$= P^{TT}D^TP^T$$

$$= PDP^T = A$$

# Orthogonal Diagonalizability and Symmetry

**Theorem.** A matrix is orthogonally diagonalizable if and only if it is symmetric.

*(You won't need to construct an orthogonal diagonalization, we'll just use NumPy.)*

# Quadratic Forms

$Q(x_1, x_2, x_3) = 3x_1^2 + 4x_2^2 - 5x_1x_2 - 10x_1x_3$

# Quadratic Forms

Nonexample. $Q(x_1) = x_1^2 + 5$

Non-example: $Q(x_1, x_2) = 5x_1^2 + x_1x_2 + x_2$ (linear term.)

**Definition.** A **quadratic form** is an function of variables $x_1, \ldots, x_n$ in which every term has degree two:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} x_i x_j$$

$x_1^2 \quad x_2^2$

*Quadratic forms are the quadratic versions the left-hand-sides of linear equations.*

# Examples

$$Q(\vec{x}) = \langle x, x \rangle$$

$$= \sum_{i=1}^{n} x_i^2$$

$$Q(\vec{x}) = \langle x, Ax \rangle$$

$$= \sum_{i=1}^{n} x_i (Ax)_i$$

$$\sum_{i=1}^{n} x_i \sum_{j=1}^{n} A_{ij} x_j$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

# Quadratic Forms and Symmetric Matrices

**Fact.** Every quadratic form can be represented as

$$\mathbf{x}^T A \mathbf{x} \qquad \langle x, Ax \rangle$$

where $A$ is <u>symmetric</u>.

Example: $5x_1^2 + 7x_2^2$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 5x_1 \\ 7x_2 \end{bmatrix} = 5x_1^2 + 7x_2^2$$

# Example: Computing the Quadratic Form for a Matrix

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

$$\begin{bmatrix} (3x_1 - 2x_2) & (-2x_1 + 7x_2) \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$$

This means, given a symmetric matrix $A$, we can compute its corresponding quadratic form:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{bmatrix}$$

$$= x_1(3x_1 - 2x_2) + x_2(-2x_1 + 7x_2) = 3x_1^2 - 2x_1x_2 - 2x_2x_1 + 7x_2^2$$

$$= 3x_1^2 - 4x_1x_2 + 7x_2^2$$

# Quadratic forms and Symmetric Matrices (Again)

Furthermore, we can generally say

$$\mathbf{x}^T A\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

Verify:

# A Slightly more Complicated Example

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & 0 \\ -1 & 0 & 5 \end{bmatrix}$$

Let's expand $\mathbf{x}^T A \mathbf{x}$:

# Matrices from Quadratic Forms

$$Q(\mathbf{x}) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$$

We can also go in the other direction. Let's express this as $\mathbf{x}^T A\mathbf{x}$:

# How To: Matrices of Quadratic Forms

**Problem.** Given a quadratic form $Q(\mathbf{x})$, find the symmetric matrix $A$ such that $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$.

**Solution.**

» if $Q(\mathbf{x})$ has the term $\alpha x_i^2$ then $A_{ii} = \alpha$

» if $Q(\mathbf{x})$ has the term $\alpha x_i x_j$, then $A_{ij} = A_{ji} = \dfrac{\alpha}{2}$

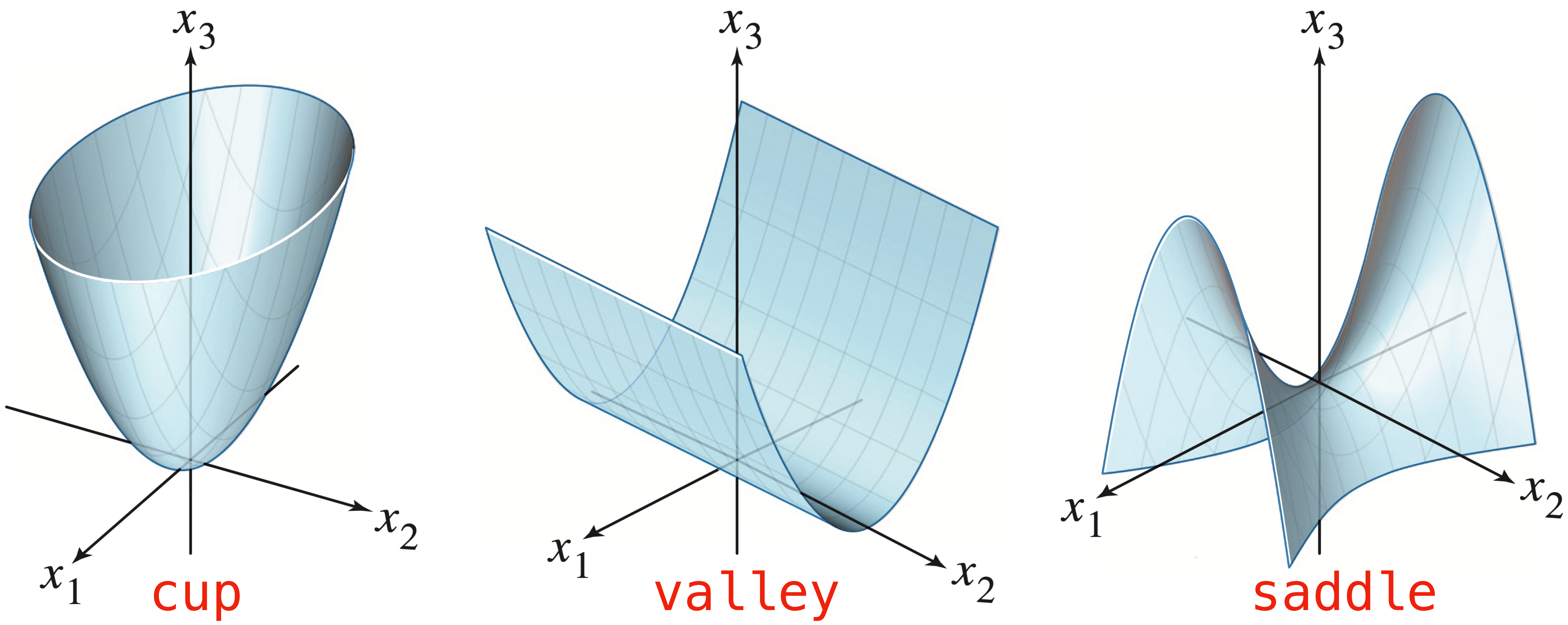
# Question

$$Q(x_1, x_2, x_3, x_4) = x_1^2 + 3x_2^2 - 2x_3x_4 - 6x_4^2 + 7x_1x_3$$

*Find the symmetric matrix* $A$ *such that* $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$*.*

# Answer

# Shapes of of Quadratic Forms in $\mathbb{R}^3$

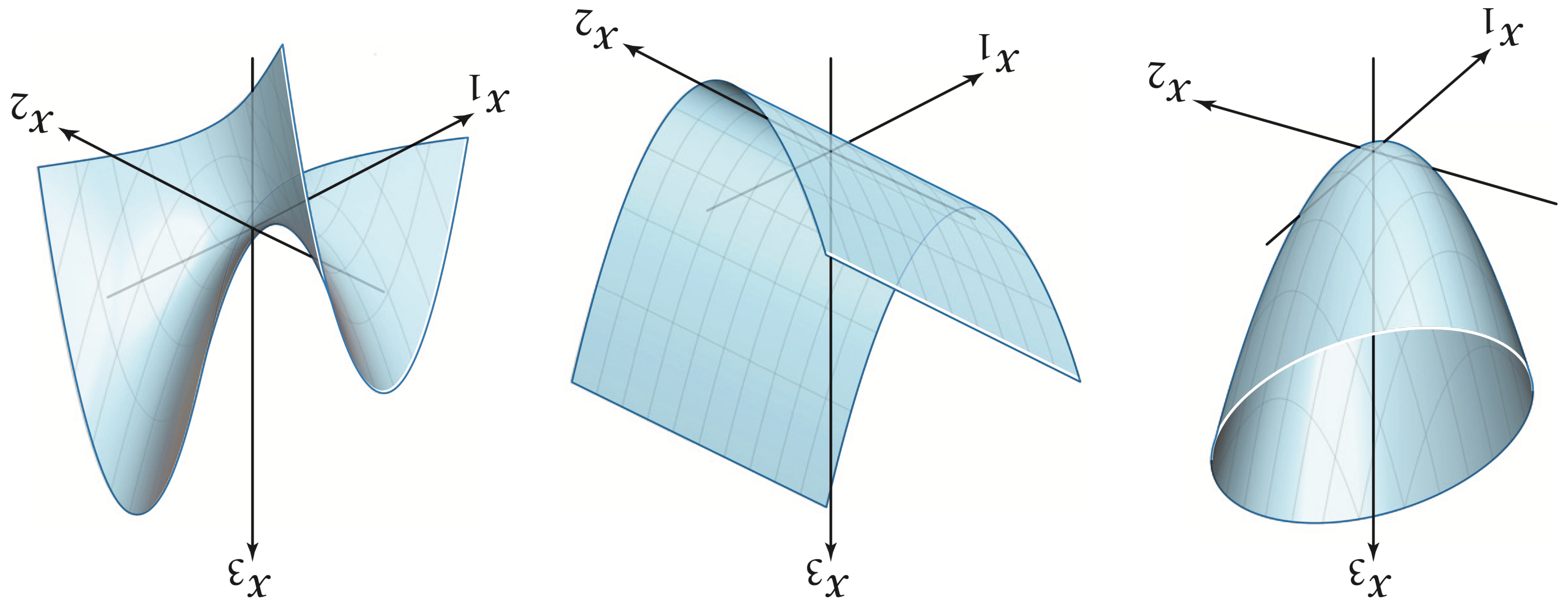



There are essentially three possible shapes (six if you include the negations).

*Can we determine what shape it will be mathematically?*

Linear Algebra and its Applications, Lay, Lay, McDonald

# Shapes of of Quadratic Forms in $\mathbb{R}^3$

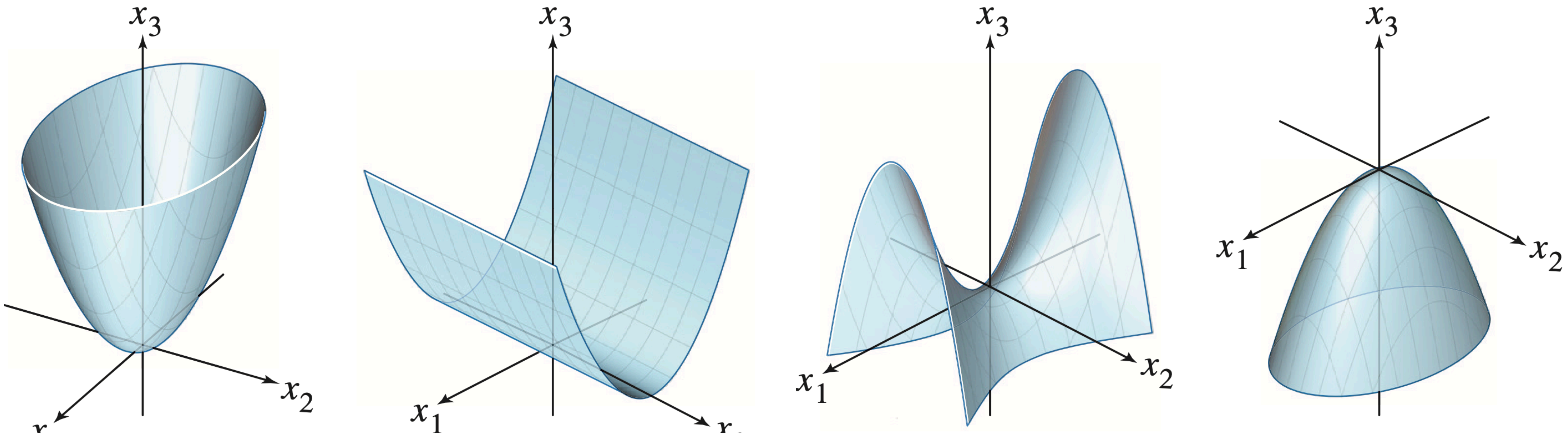



There are essentially three possible shapes (six if you include the negations).

*Can we determine what shape it will be mathematically?*

Linear Algebra and its Applications, Lay, Lay, McDonald

# Definiteness



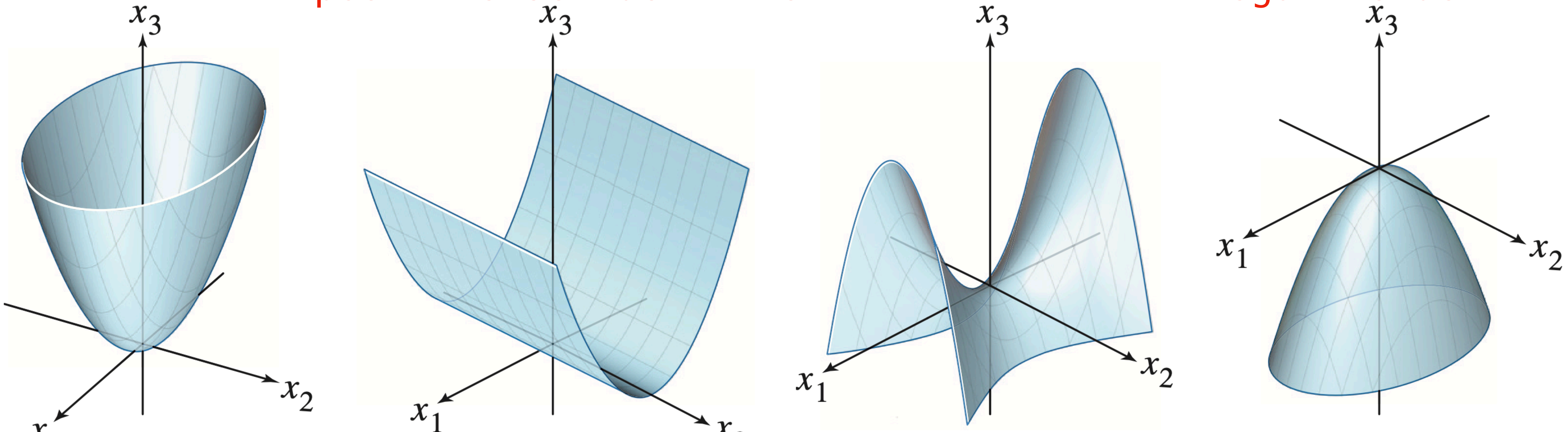

$Q(\mathbf{x}) > 0$ $Q(\mathbf{x}) \geq 0$ $Q(\mathbf{x})$ can be + & – $Q(\mathbf{x}) < 0$

For $\mathbf{x} \neq \mathbf{0}$, each of the above graphs satisfy the associated properties.

# Definiteness

positive semidefinite

negative definite




$Q(\mathbf{x}) > 0$ | $Q(\mathbf{x}) \geq 0$ | $Q(\mathbf{x})$ can be + & – | $Q(\mathbf{x}) < 0$

positive definite

indefinite

For $\mathbf{x} \neq \mathbf{0}$, each of the above graphs satisfy the associated properties.

# Definiteness and Eigenvectors

**Theorem.** For a symmetric matrix $A$, the quadratic form $\mathbf{x}^T A \mathbf{x}$

- **positive definite** ≡ all positive eigenvalues
- **positive semidefinite** ≡ all nonnegative eigenvalues
- **indefinite** ≡ positive and negative eigenvalues
- **negative definite** ≡ all negative eigenvalues

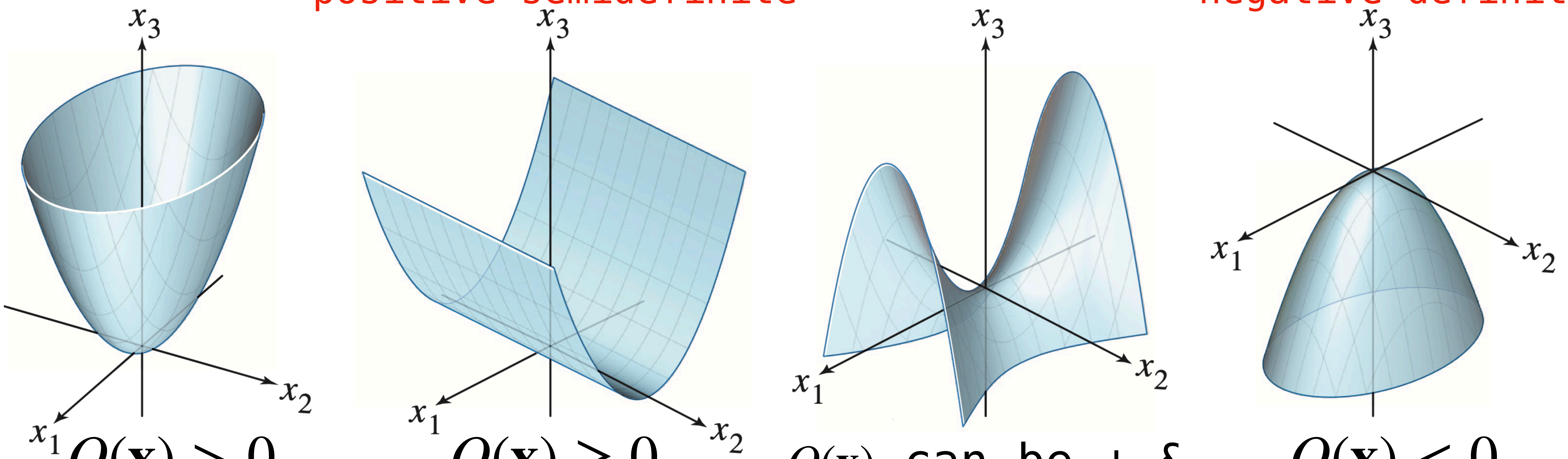
# Definiteness

all nonneg. eigenvals

positive semidefinite

all neg. eigenvals

negative definite




$Q(\mathbf{x}) > 0$ $\qquad Q(\mathbf{x}) \geq 0$ $\qquad Q(\mathbf{x})$ can be + & − $\qquad Q(\mathbf{x}) < 0$

positive definite

all pos. eigenvals

indefinite

pos. and neg. eigenvals

# Positive Definite Case

```
Let's think why this is for the positive definite case:
```

# Example

$$Q(x_1, x_2, x_3) = 3x_1^2 + x_2^2 + 4x_2x_3 + x_3^2$$

```
Let's determine which case this is:
```

# Constrained Optimization

# In General

# In General

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a set of vectors $X$ from $\mathbb{R}^n$ the **constrained minimization problem** for $f$ over $X$ is the problem of determining

$$\min_{\mathbf{v} \in X} f(\mathbf{v})$$

# In General

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a set of vectors $X$ from $\mathbb{R}^n$ the **constrained minimization problem** for $f$ over $X$ is the problem of determining

$$\min_{\mathbf{v} \in X} f(\mathbf{v})$$

(analogously for maximization)

# In General

Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a set of vectors $X$ from $\mathbb{R}^n$ the **constrained minimization problem** for $f$ over $X$ is the problem of determining

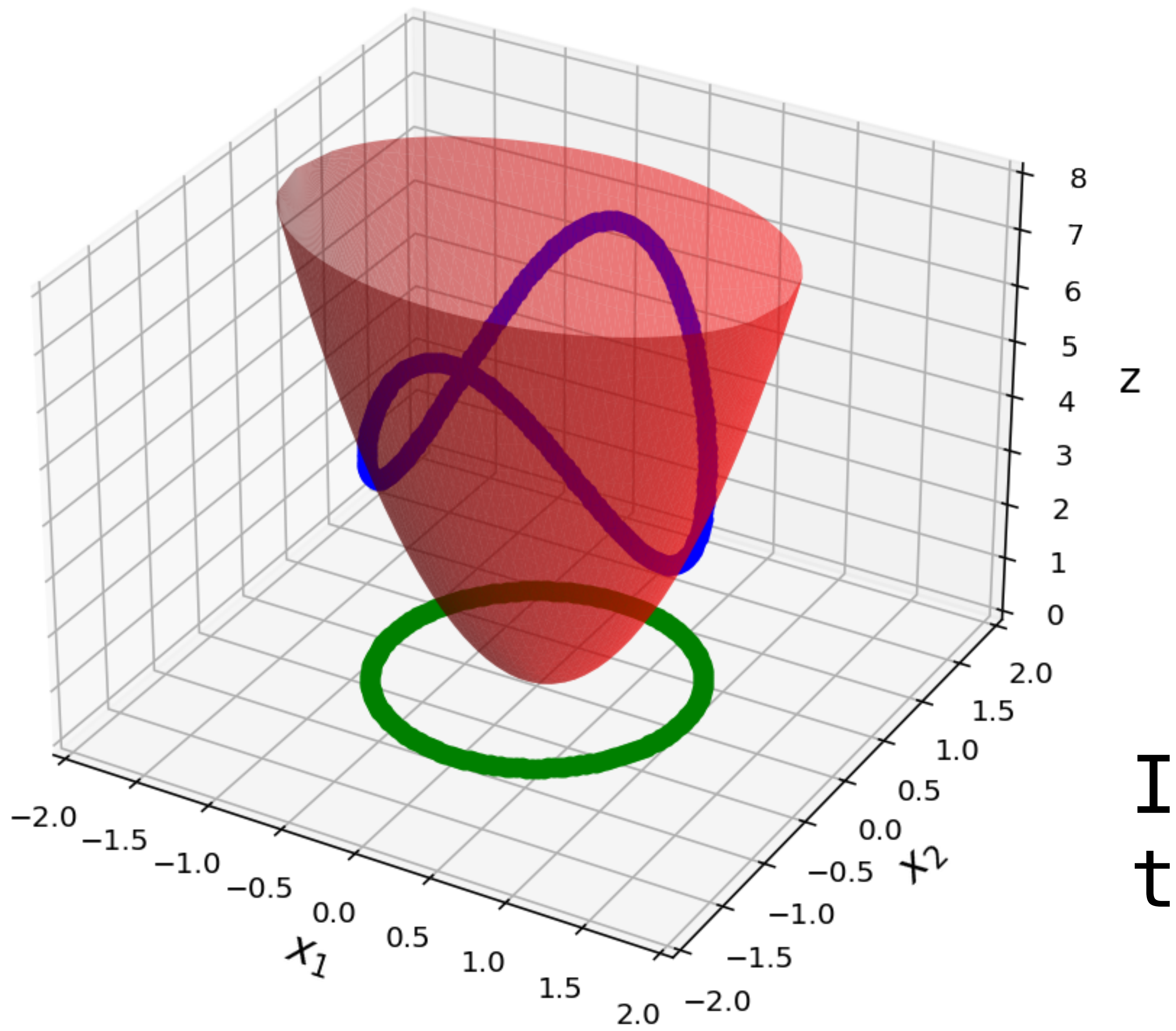
$$\min_{\mathbf{v} \in X} f(\mathbf{v})$$

(analogously for maximization)

*Find the smallest value of $f(\mathbf{v})$ subject to choosing a vector in $X$*

# Constrained Optimization for Quadratic Forms and Unit Vectors

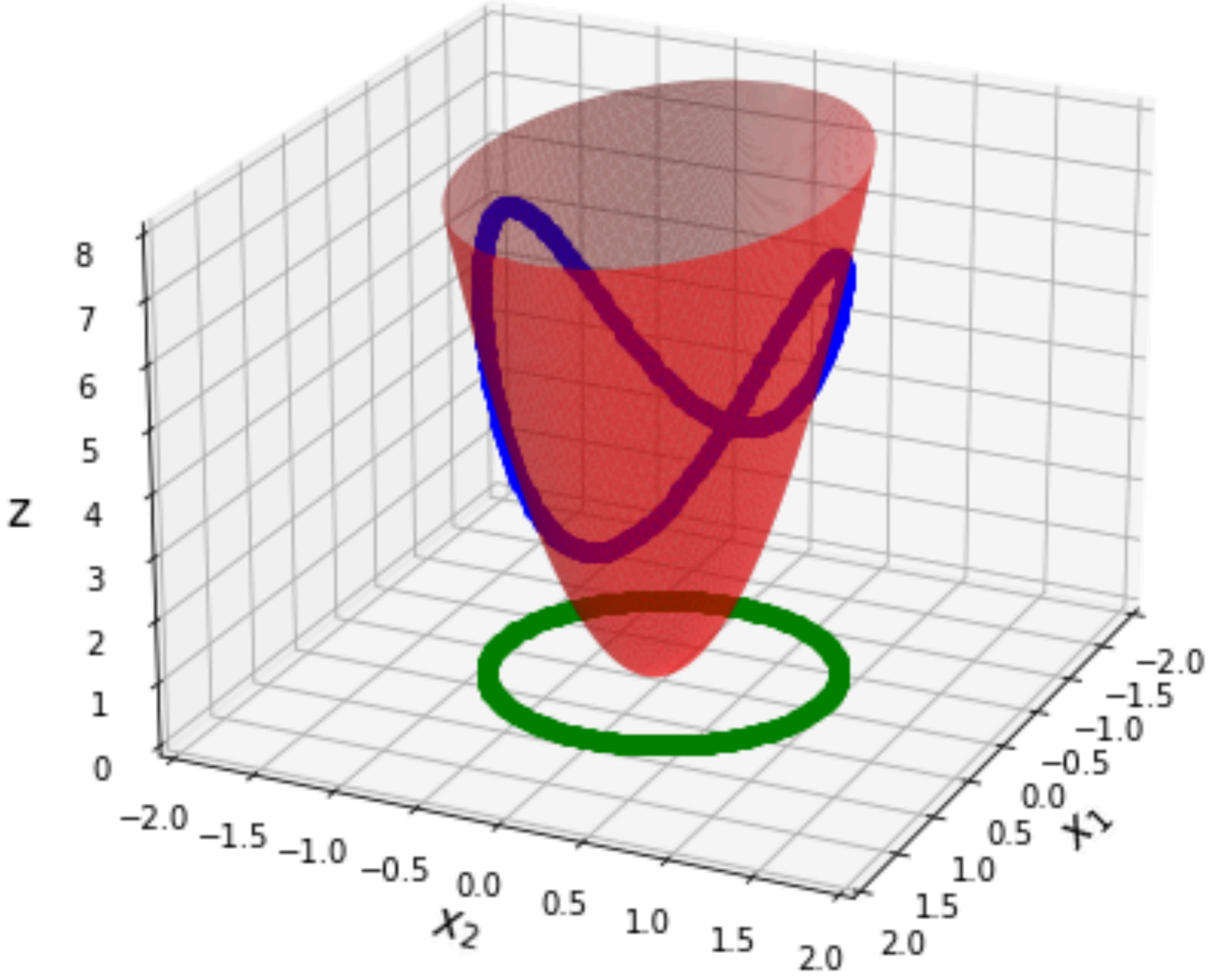
$$\text{mini/maximize } \mathbf{x}^T A \mathbf{x} \text{ subject to } \|\mathbf{x}\| = 1$$




It's common to constraint to unit vectors.

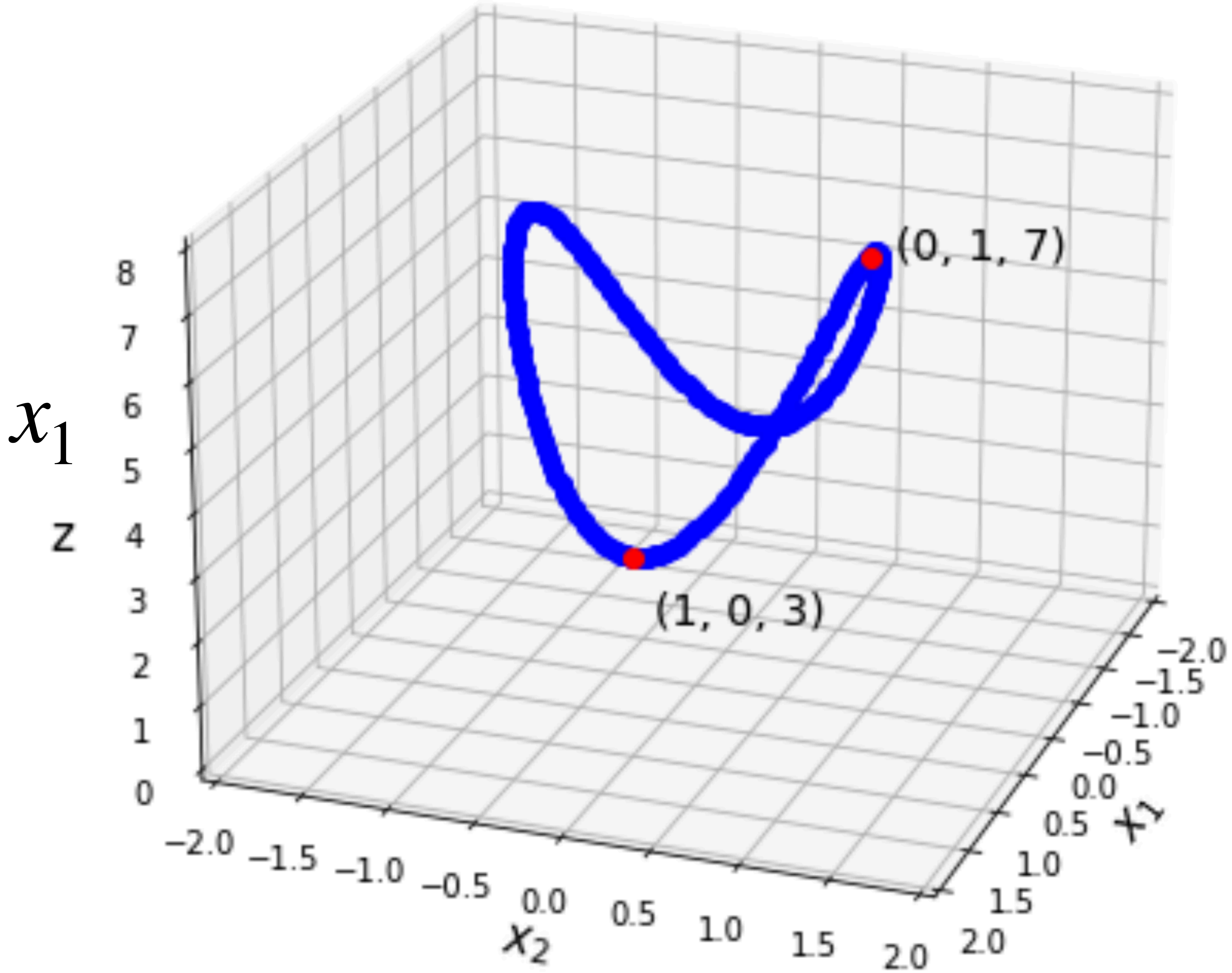
**Example:** $3x_1^2 + 7x_2^2$

`What are the min/max values?:`



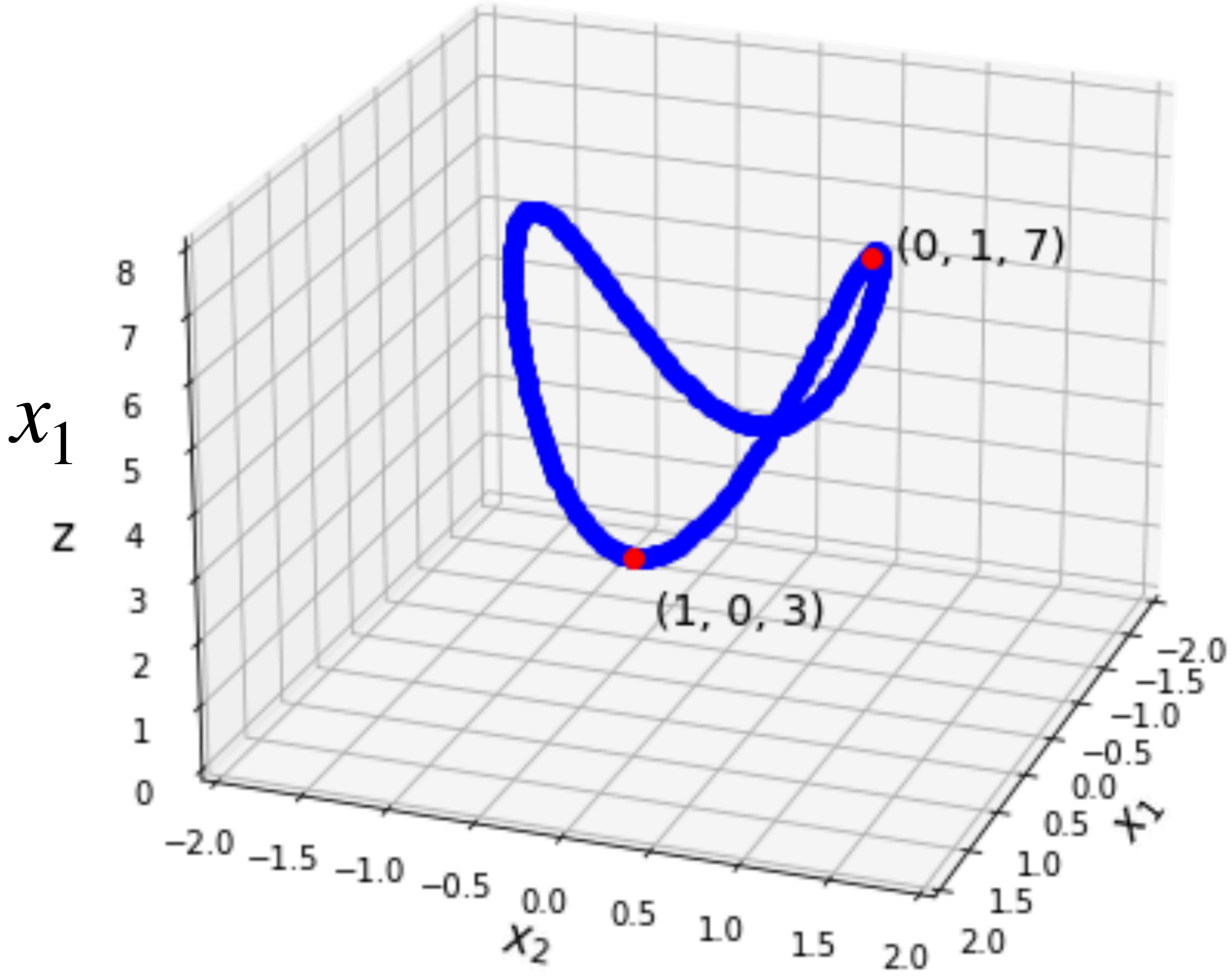
# **Example:** $3x_1^2 + 7x_2^2$

The minimum and maximum values are attained when the "weight" of the vector is distributed all on $x_1$ or $x_2$.



## **Example:** $3x_1^2 + 7x_2^2$

What is the matrix?:



# Constrained Optimization and Eigenvalues

**Theorem.** For a symmetric matrix $A$, with *largest* eigenvalue $\lambda_1$ and *smallest* eigenvalue $\lambda_n$

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} = \lambda_1 \qquad \min_{\|\mathbf{x}\|=1} \mathbf{x}^T A \mathbf{x} = \lambda_n$$

*No matter the shape of $A$, this will hold.*

# How To: Constrained Optimization

# How To: Constrained Optimization

**Problem.** Find the maximum value of $\mathbf{x}^T A \mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.

# How To: Constrained Optimization

**Problem.** Find the maximum value of $\mathbf{x}^T A \mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.

**Solution.** Find the largest eigenvalue of $A$, this will be the maximum value.

# How To: Constrained Optimization

**Problem.** Find the maximum value of $\mathbf{x}^T A \mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.

**Solution.** Find the largest eigenvalue of $A$, this will be the maximum value.

*(Use NumPy)*

# Summary

We can build models which are nonlinear functions if those functions are linear in their parameters.

We can solve constrained optimization problems using eigenvalues.