Module 1 | Lesson 1 (Part 1)

The Squared Error Criterion and the Method of Least Squares

Module 1 | Least Squares

In this module

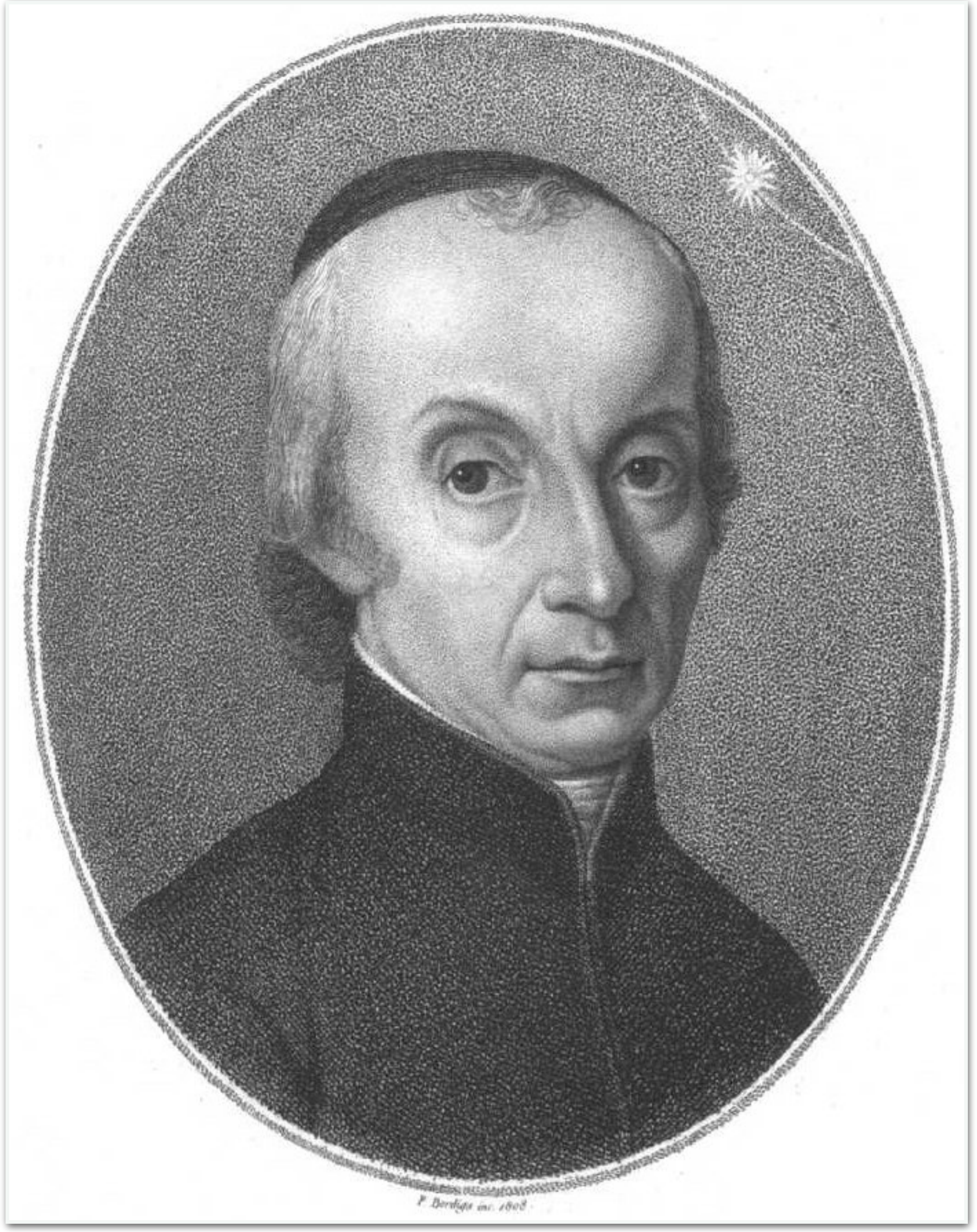
- The history of the method of *least squares*
- Ordinary and weighted least squares
- Recursive least squares
- Maximum likelihood and the method of least squares

The Squared Error Criterion and the Method of Least Squares

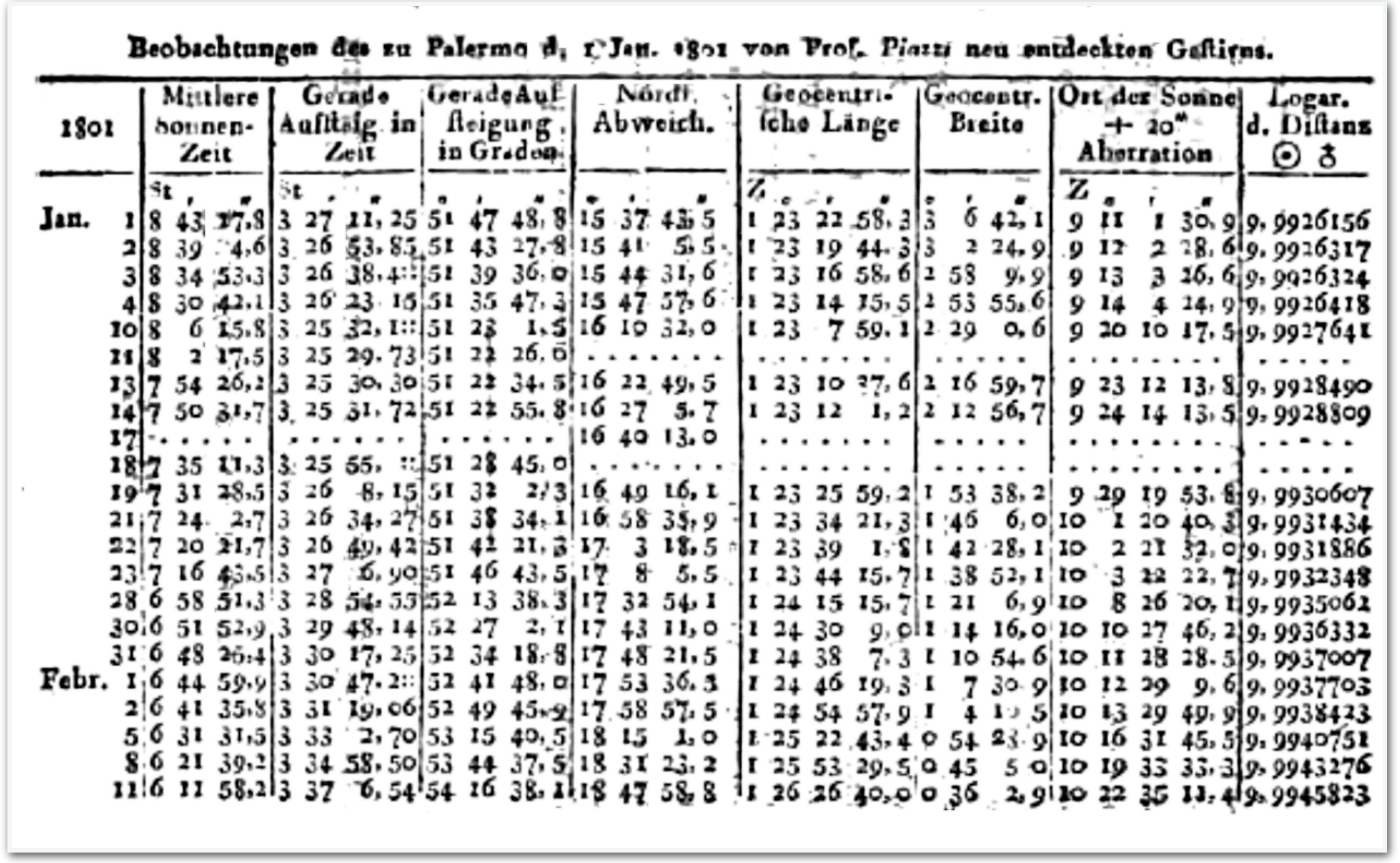
By the end of this video, you will be able to…

- Describe how the method of least squares was used in the discovery of Ceres
- Describe the least error criterion and how it's used in parameter estimation
- Derive the normal equations for least squares parameter estimation

Giuseppe Piazzi and Ceres



Giuseppe Piazzi



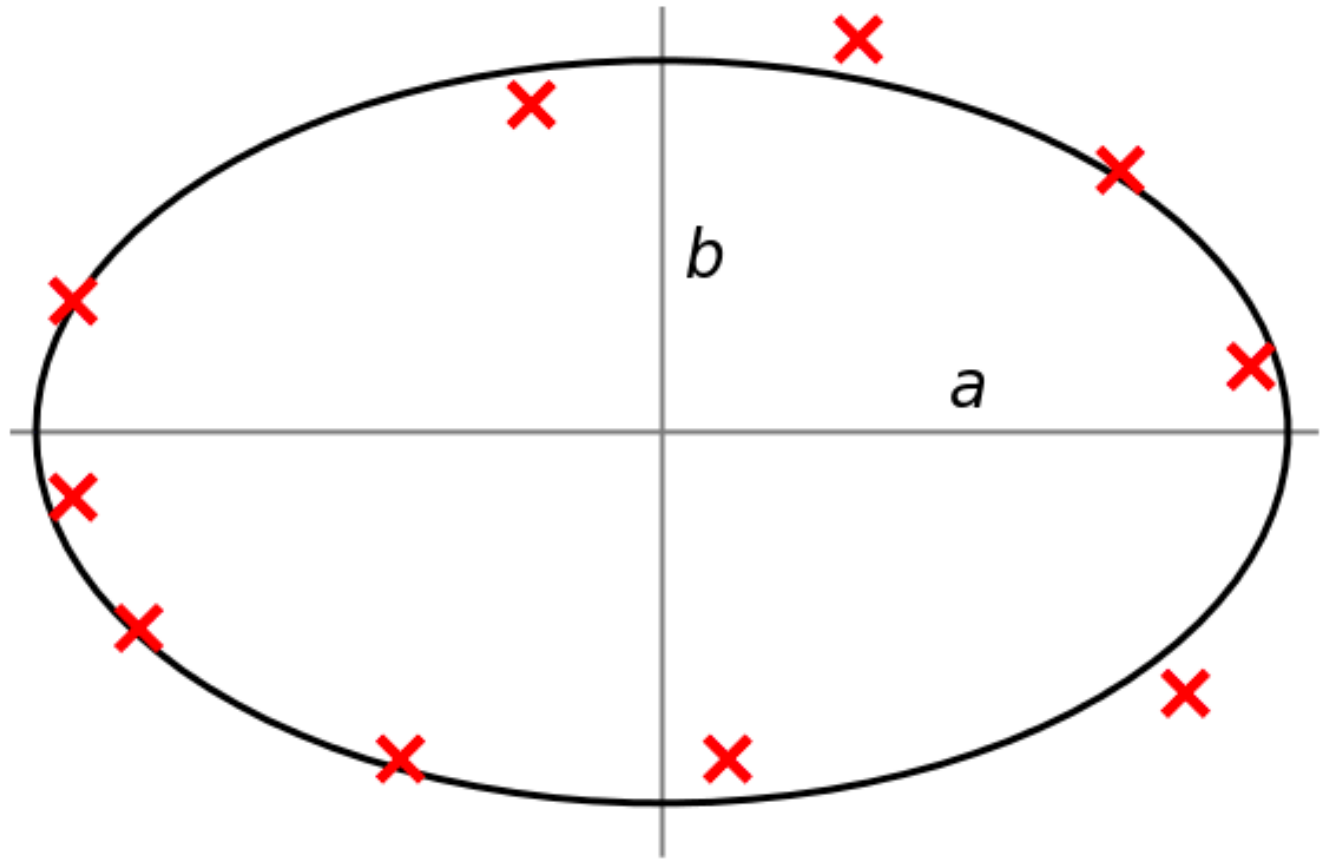
Beobachtungen des zu Palermo d. 1. Jan. 1801 von Prof. Piazzi neu entdeckten Gestirns.

1801		Mittlere Sonnen-Zeit	Gerade Aufstaig in Zeit	Gerade Aufsteigung in Graden	Nördl. Abweich.	Geocentrische Länge	Geocentr. Breite	Ort der Sonne + 20" Aberration	Logar. d. Distanz ☉ ♂
		St ′ ″	St ′ ″	° ′ ″	° ′ ″	Z ° ′ ″	° ′ ″	Z ° ′ ″	
Jan.	1	8 43 27,8	3 27 11,25	51 47 48,8	15 37 43,5	1 23 22 58,3	3 6 42,1	9 11 1 30,9	9,9926156
	2	8 39 4,6	3 26 53,85	51 43 27,8	15 41 5,5	1 23 19 44,3	3 2 24,9	9 12 2 28,6	9,9926317
	3	8 34 53,3	3 26 38,4::	51 39 36,0	15 44 31,6	1 23 16 58,6	2 58 9,9	9 13 3 26,6	9,9926324
	4	8 30 42,1	3 26 23,16	51 35 47,3	15 47 57,6	1 23 14 15,5	2 53 55,6	9 14 4 24,9	9,9926418
	10	8 6 15,8	3 25 32,1::	51 23 1,5	16 10 32,0	1 23 7 59,1	2 29 0,6	9 20 10 17,5	9,9927641
	11	8 2 17,5	3 25 29,73	51 22 26,0	. . . .	. . . .	. . . . .	. . . .	. . . . .
	13	7 54 26,2	3 25 30,30	51 22 34,5	16 22 49,5	1 23 10 27,6	2 16 59,7	9 23 12 13,8	9,9928490
	14	7 50 31,7	3 25 31,72	51 22 55,8	16 27 5,7	1 23 12 1,2	2 12 56,7	9 24 14 13,5	9,9928809
	17	. . . . . .	. . . . . .	. . . . . .	16 40 13,0	. . . .	. . . . .	. . . .	. . . . .
	18	7 35 11,3	3 25 55,::	51 28 45,0	. . . .	. . . .	. . . . .	. . . .	. . . . .
	19	7 31 28,5	3 26 8,15	51 32 2,3	16 49 16,1	1 23 25 59,2	1 53 38,2	9 29 19 53,8	9,9930607
	21	7 24 2,7	3 26 34,27	51 38 34,1	16 58 35,9	1 23 34 21,3	1 46 6,0	10 1 20 40,3	9,9931434
	22	7 20 21,7	3 26 49,42	51 42 21,3	17 3 18,5	1 23 39 1,8	1 42 28,1	10 2 21 32,0	9,9931886
	23	7 16 43,5	3 27 6,90	51 46 43,5	17 8 5,5	1 23 44 15,7	1 38 52,1	10 3 22 22,7	9,9932348
	28	6 58 51,3	3 28 54,55	52 13 38,3	17 32 54,1	1 24 15 15,7	1 21 6,9	10 8 26 20,1	9,9935062
	30	6 51 52,9	3 29 48,14	52 27 2,1	17 43 11,0	1 24 30 9,0	1 14 16,0	10 10 27 46,2	9,9936332
	31	6 48 26,4	3 30 17,25	52 34 18,8	17 48 21,5	1 24 38 7,3	1 10 54,6	10 11 28 28,5	9,9937007
Febr.	1	6 44 59,9	3 30 47,2::	52 41 48,0	17 53 36,3	1 24 46 19,3	1 7 30,9	10 12 29 9,6	9,9937703
	2	6 41 35,8	3 31 19,06	52 49 45,9	17 58 57,5	1 24 54 57,9	1 4 1,5	10 13 29 49,9	9,9938423
	5	6 31 31,5	3 33 2,70	53 15 40,5	18 15 1,0	1 25 22 43,4	0 54 23,9	10 16 31 45,5	9,9940751
	8	6 21 39,2	3 34 58,50	53 44 37,5	18 31 23,2	1 25 53 29,5	0 45 5,0	10 19 33 33,3	9,9943276
	11	6 11 58,2	3 37 6,54	54 16 38,1	18 47 58,8	1 26 26 40,0	0 36 2,9	10 22 35 11,4	9,9945823

Piazzi's 24 observations

Karl Friedrich Gauss

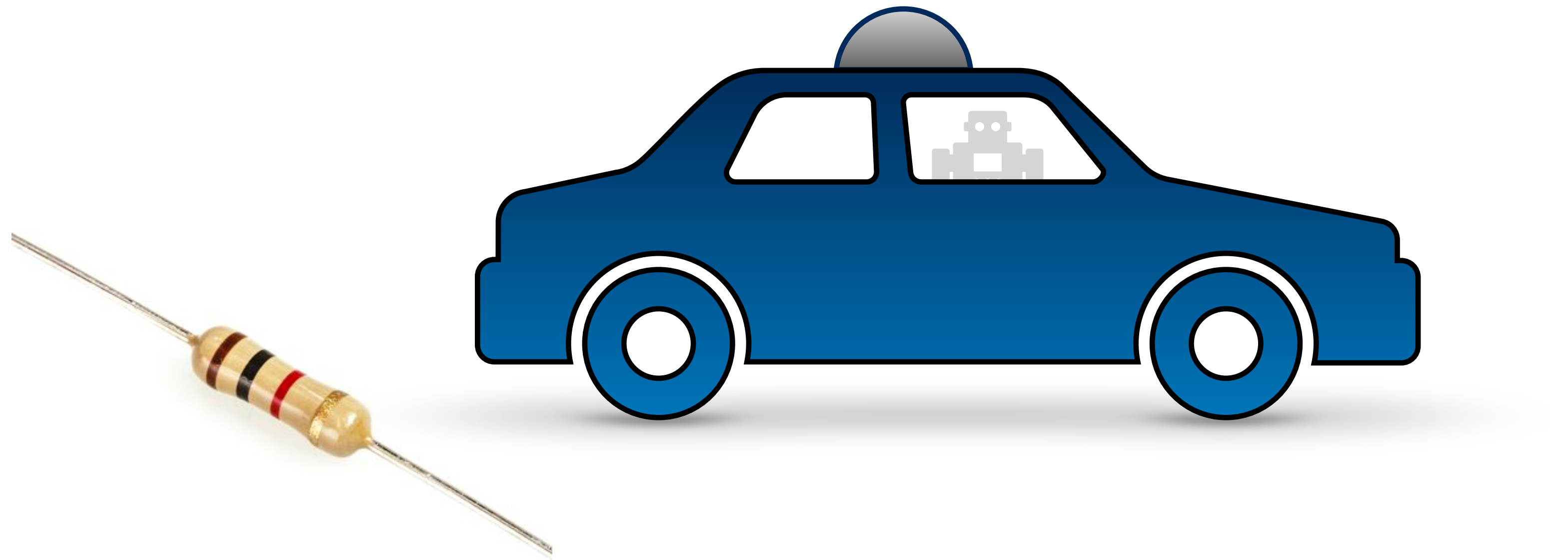
Karl Friedrich Gauss
'Princeps mathematicorum'



*Gauss used the **method of least squares** to determine the orbital parameters of Ceres.*

Least Squares

> *The **most probable value** of the unknown quantities will be that in which the **sum of the squares** of the differences between the actually observed and the computed values multiplied by numbers that measure the degree of precision is a **minimum**.*
>
> \- Karl Friedrich Gauss



Resistor in the drive-system of a car

Multimeter

Estimating Resistance

Measurement	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

R?

Estimating Resistance

Measurement	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Let x be the resistance. Assume it is a **constant**, but **unknown**.

We make measurements, y, of the resistance.

We model our measurements as corrupted by noise v.

$$\boxed{y = x + v}$$

Estimating Resistance

Measurement	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Measurement Model

'Actual' resistance *Measurement noise*

$$y_1 = x + v_1$$

$$y_2 = x + v_2$$

$$y_3 = x + v_3$$

$$y_4 = x + v_4$$

Estimating Resistance

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Measurement Model

$$y_1 = x + v_1$$

$$y_2 = x + v_2$$

$$y_3 = x + v_3$$

$$y_4 = x + v_4$$

Squared Error

$$e_1^2 = (y_1 - x)^2$$

$$e_2^2 = (y_2 - x)^2$$

$$e_3^2 = (y_3 - x)^2$$

$$e_4^2 = (y_4 - x)^2$$

The squared error *criterion*:

$$\hat{x}_{\mathrm{LS}} = \mathrm{argmin}_{\,x} \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\mathrm{LS}}(x)$$

The 'best' estimate of resistance is the one that minimizes the sum of squared errors

Minimizing the Squared Error Criterion

$$\boxed{\hat{x}_{\mathrm{LS}} = \operatorname{argmin}_x \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\mathrm{LS}}(x)}$$

Let's re-write our criterion using vectors:

This matrix is called the ***'Jacobian'***

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \mathbf{y} - \mathbf{H}x$$

$$= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x$$

Minimizing the Squared Error Criterion

$$\boxed{\hat{x}_{\mathrm{LS}} = \operatorname{argmin}_x \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\mathrm{LS}}(x)}$$

Now, we can express our criterion as follows,

$$\begin{aligned}\mathscr{L}_{\mathrm{LS}}(x) = e_1^2 + e_2^2 + e_3^2 + e_4^2 &= \mathbf{e}^T\mathbf{e} \\ &= (\mathbf{y} - \mathbf{H}x)^T(\mathbf{y} - \mathbf{H}x) \\ &= \mathbf{y}^T\mathbf{y} - x^T\mathbf{H}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}x + x^T\mathbf{H}^T\mathbf{H}x\end{aligned}$$

Minimizing the Squared Error Criterion

$$\boxed{\mathscr{L}(x) = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T\mathbf{y} - x^T\mathbf{H}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}x + x^T\mathbf{H}^T\mathbf{H}x}$$

To minimize this, we can compute the partial derivative with respect to our parameter, set to 0, and solve for an extremum:

$$\left.\frac{\partial \mathscr{L}}{\partial x}\right|_{x=\hat{x}} = -\mathbf{y}^T\mathbf{H} - \mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0$$

$$-2\mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0$$

Re-arranging, we arrive at:

$$\hat{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$$

x-hat minimizes our squared error criterion!

Minimizing the Squared Error Criterion

Careful! We will only be able to solve for $\hat{x}$ if $(\mathbf{H}^T\mathbf{H})^{-1}$ exists.

If we have m measurements, and n unknown parameters, then:

$$\mathbf{H} \in \mathbb{R}^{m \times n} \qquad \mathbf{H}^T\mathbf{H} \in \mathbb{R}^{n \times n}$$

This means that $(\mathbf{H}^T\mathbf{H})^{-1}$ exists only if there are at least as many measurements as there are unknown parameters:

$$m \geq n$$

Minimizing the Squared Error Criterion

Returning to our problem, we see that:

$$\mathbf{y} = \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

$$\hat{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$$

$$= \left([1111] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} [1111] \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} = \frac{1}{4}(1068 + 988 + 1002 + 996) = 1013.5 \quad \text{Ohms}$$

The least squares solution is just the mean of our measurements!

Method of Least Squares | Assumptions

- Our measurement model, $y = x + v$, is **linear.**
- Measurements are **equally weighted.**
  (we do not suspect that some have more noise than others).

Summary | The Method of Least Squares

- Pioneered by Gauss to determine the orbit of the planetoid *Ceres*
- Least squares finds the parameters which minimize the *Least Squares Criterion*
- Ordinary least squares assumes that measurements are weighted equally, measurement model is linear