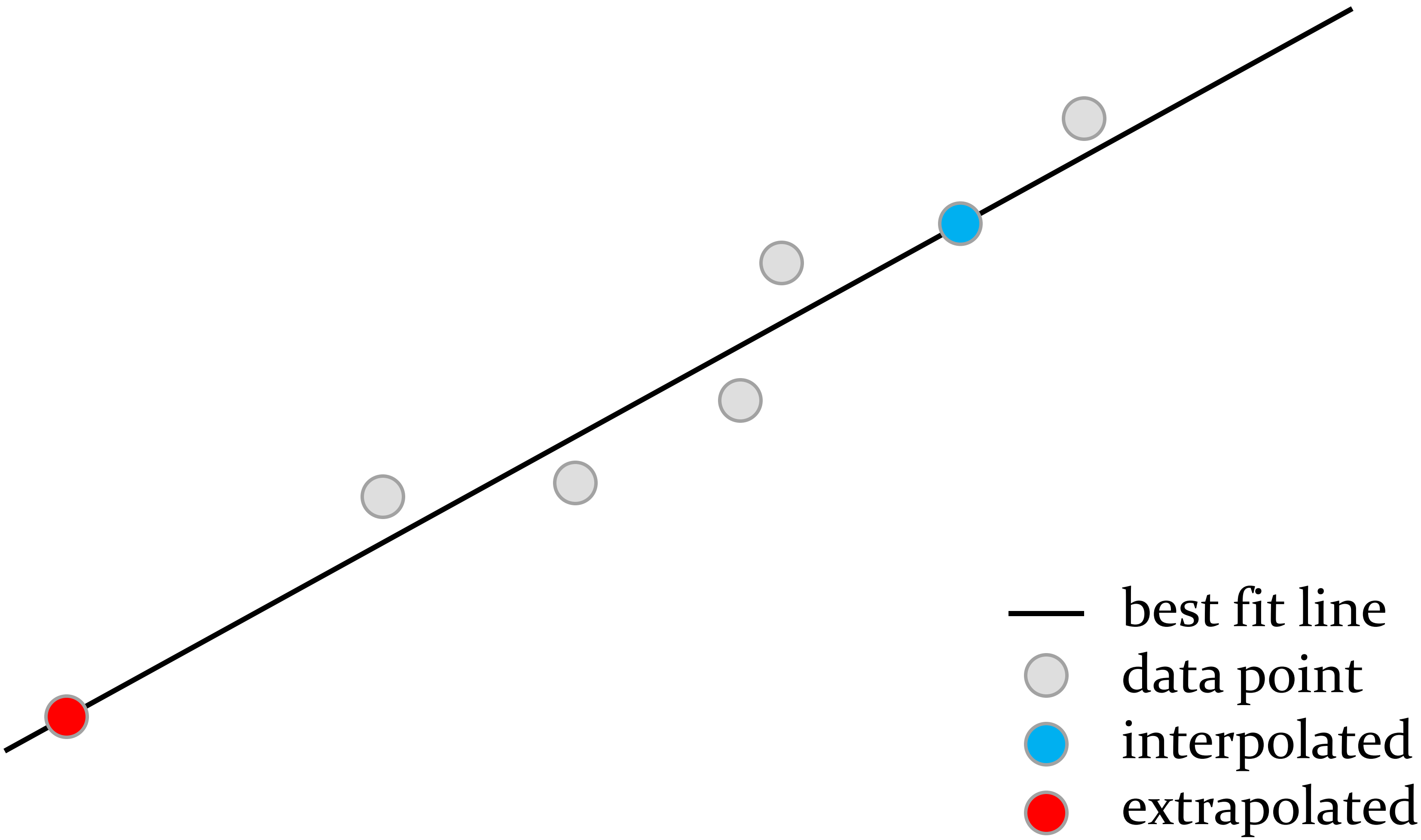
Line and curve fitting

Line and curve fitting

- a very common procedure in data analysis is fitting a line or curve to measured data points
- useful for many reasons
  - visualize measurements
  - find a functional relationship between two or more variables
  - infer results between data points (interpolation)
  - infer results outside of the range of data points (extrapolation)

Line and curve fitting



Computing the best fit line

- many undergraduate physics experiments involve linear relationships, or relationships that can be converted to linear relationships
- in such experiments you end up trying to find the best fit line to a set of measurements

$$y = a + bx$$

- the y-intercept a and slope b are typically found by software or through a pair of equations

Computing the best fit line

- for n data points (x_i, y_i)

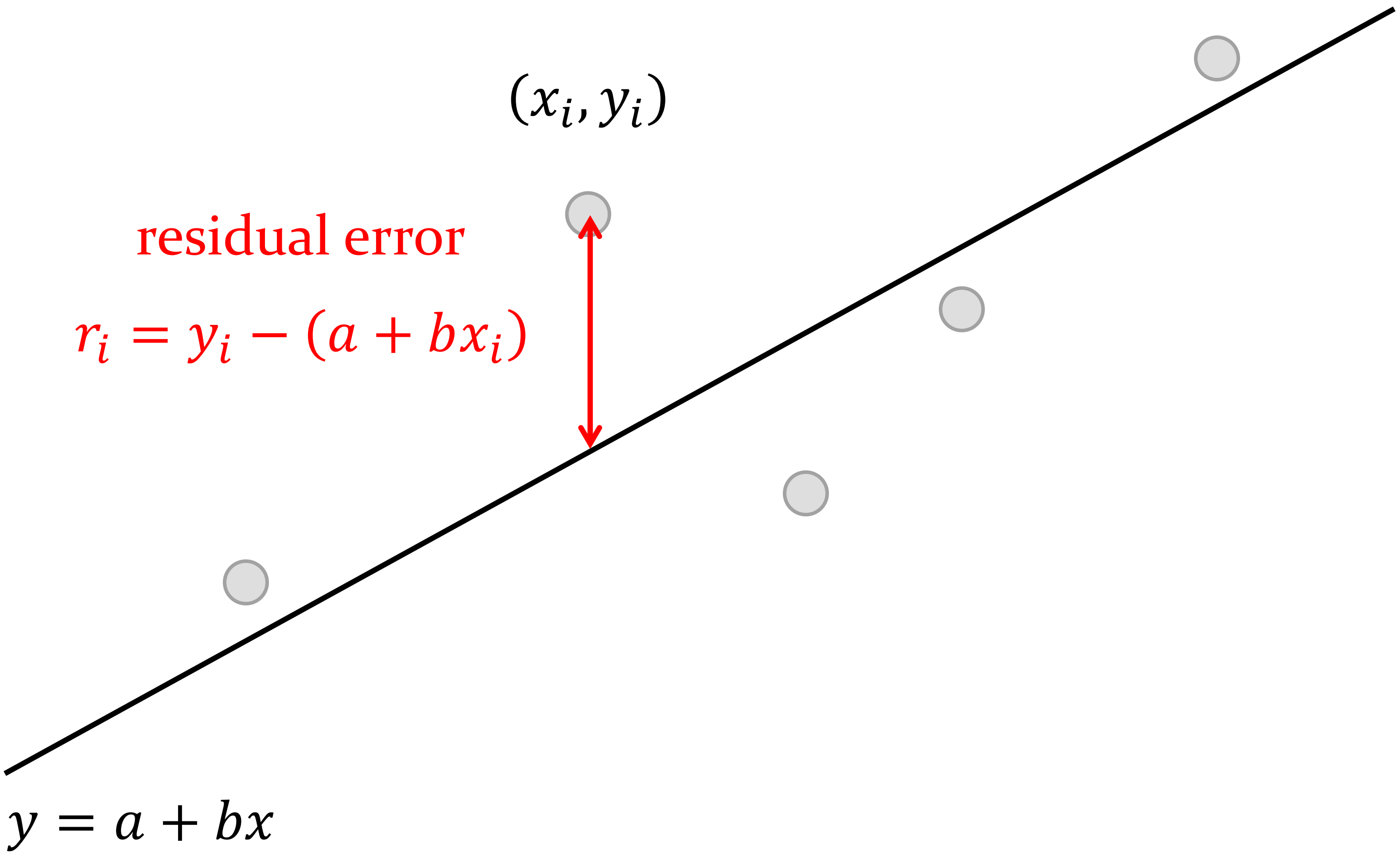
$$a = \frac{\bar{y}\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \bar{x}\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^2}$$

$$b = \frac{\left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^{\,2} - n\bar{x}^2}$$

where $\bar{x}$ and $\bar{y}$ are the average values of the x_i and y_i

- where did these come from?

Computing the best fit line



Computing the best fit line

- Gauss determined that the best fit line minimizes the sum of the squared residual errors, i.e.,
  - find a and b that minimizes

$$S = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

  - this is the principle of least-squares (Gauss-Markov theorem)

- from calculus you know that the minimum occurs when the partial derivatives are equal to zero

Computing the best fit line

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n}(y_i - (a + bx_i)) = 0$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n}(y_i - (a + bx_i))x_i = 0$$

Computing the best fit line

- a little bit of algebra yields

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2 = \sum_{i=1}^{n} x_i y_i$$

which can be written in matrix form as

$$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma {x_i}^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

Computing the best fit line

- if you compute

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma {x_i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

  you will get the equations on slide 5

- see Question 3 from Test 1-afternoon

Line fitting in MATLAB

- MATLAB provides a function for performing least-squares polynomial fitting
  - a line is a polynomial of degree 1

```
>> x = [-4; 3.7; 0; 2.5; 1.2; -2.8; -1.4];
>> y = [-37; 38; 0; 29; 21; -21; -8];
>> polyfit(x, y, 1)
ans =
    9.7436    4.2564
```