

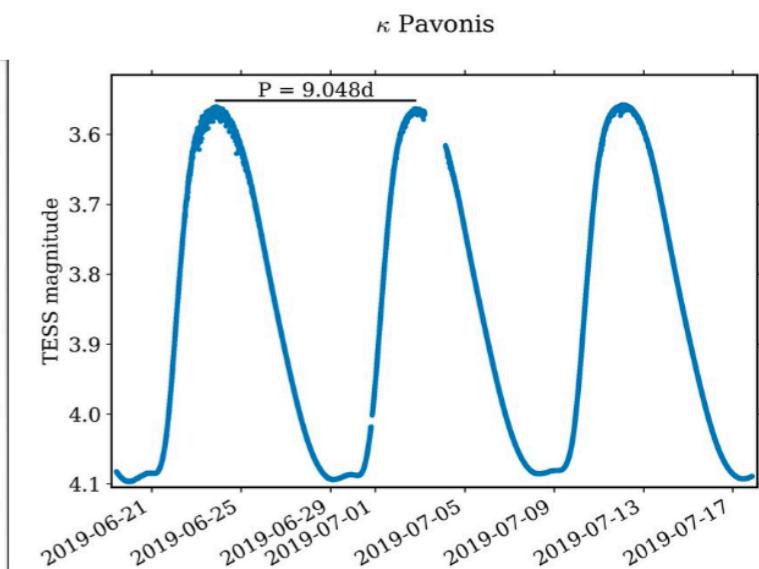
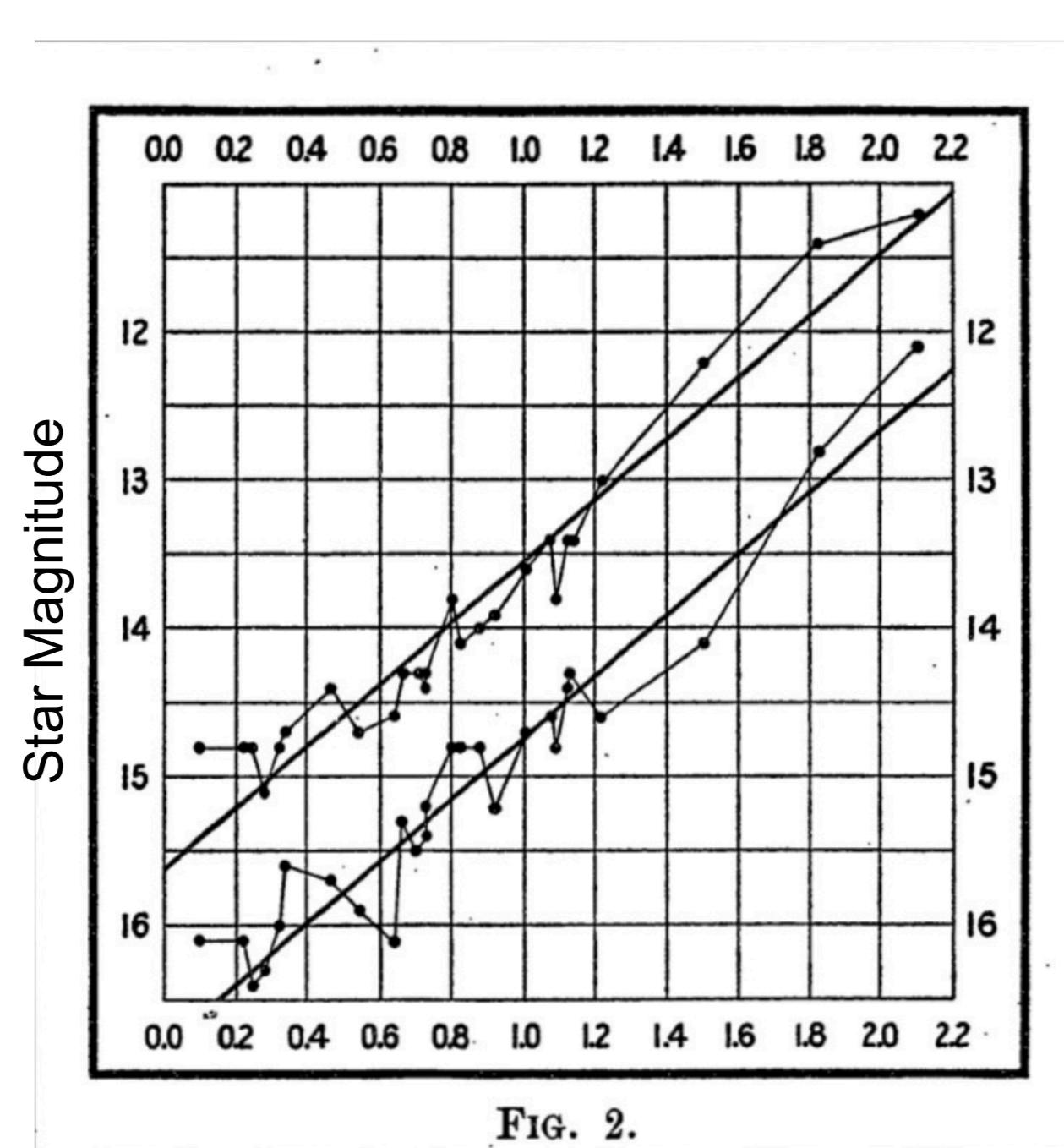
Lecture 4: Minimizing



Cambridge, Ma 1908

An Observation at Harvard

Henrietta
Leavitt,
Harvard



Cepheid variable stars oscillate in amplitude

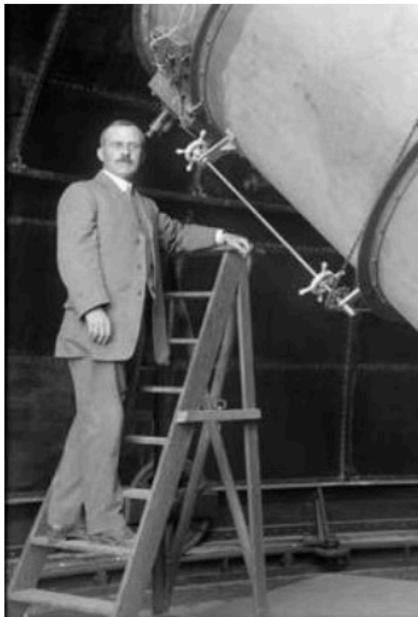


Washington, DC 1920

The great debate

The universe is
all contained in
the milky way

Heber
Curtis,
Lick (UC)

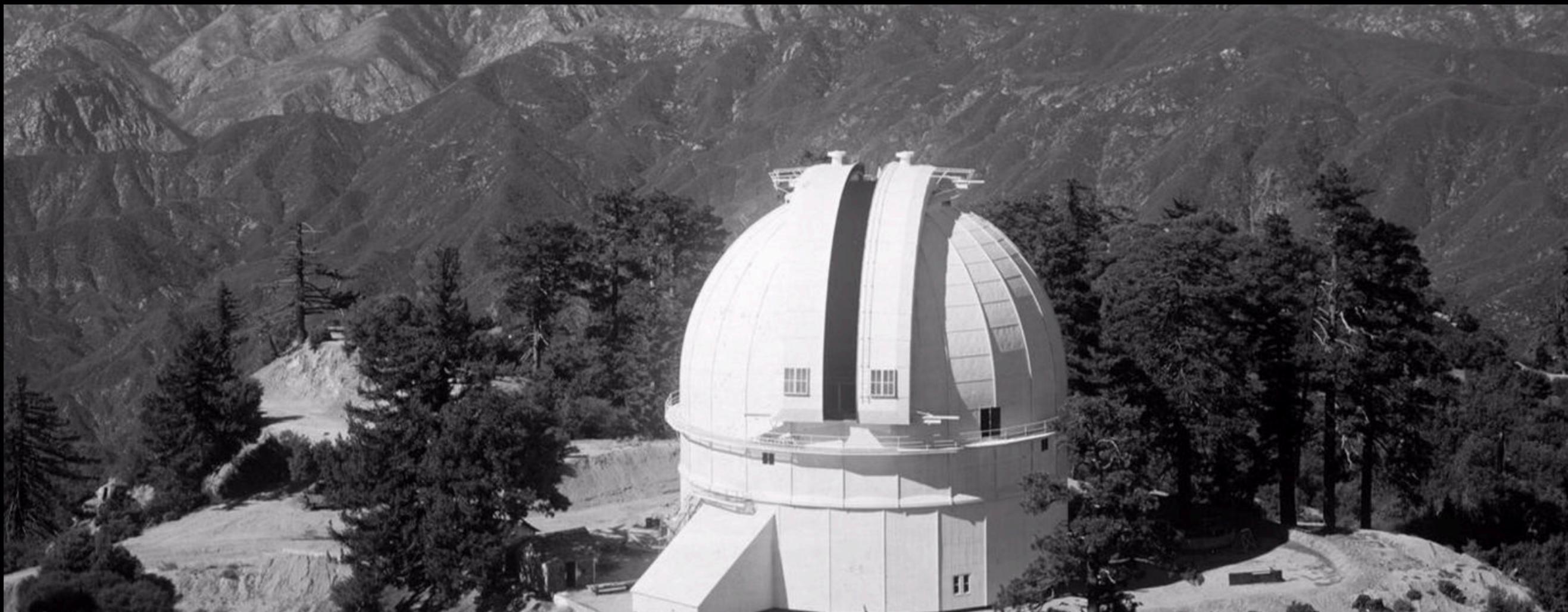


There are many
galaxies. Look at
andromeda!

Harlow
Shapely,
Harvard



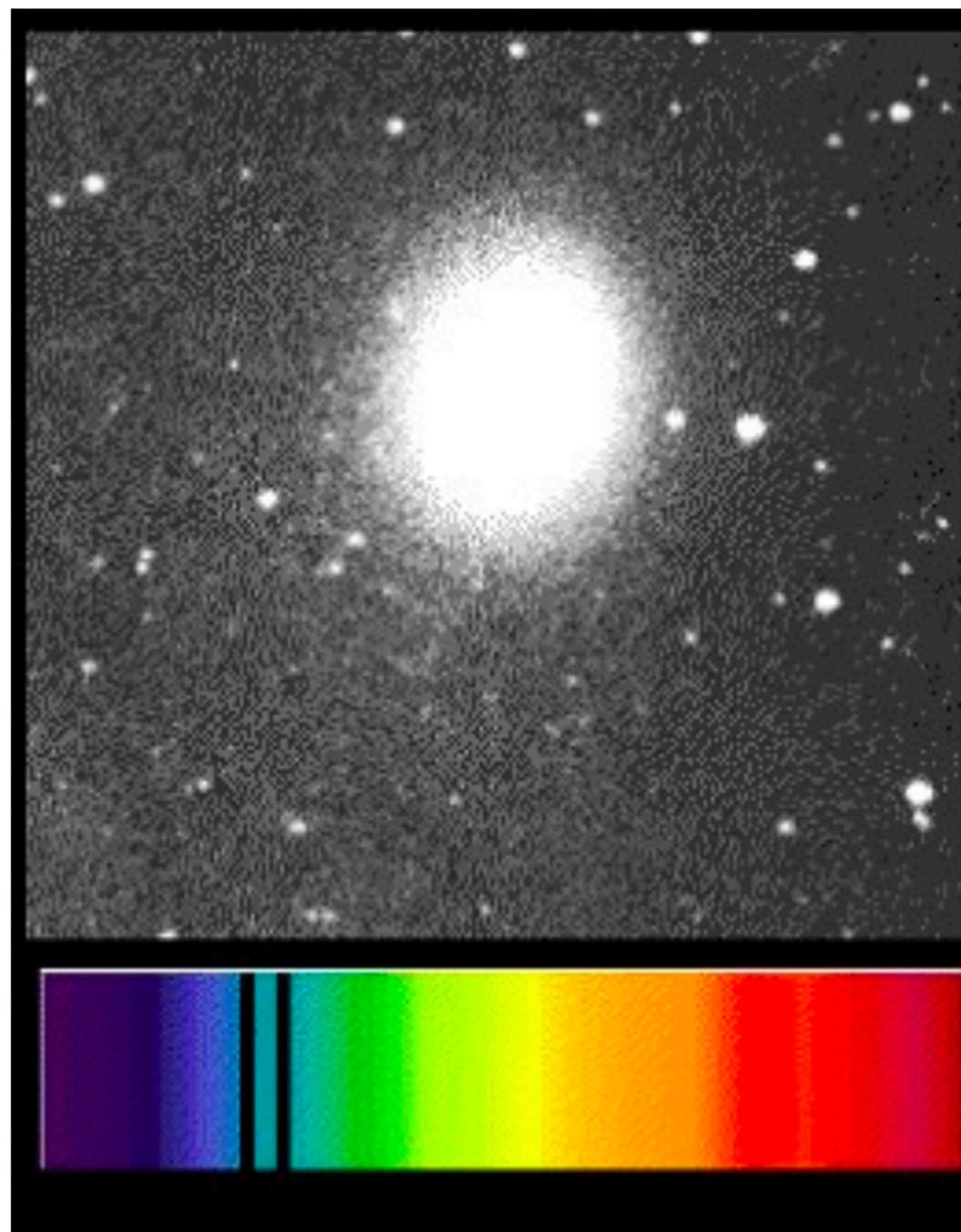
How big is the universe



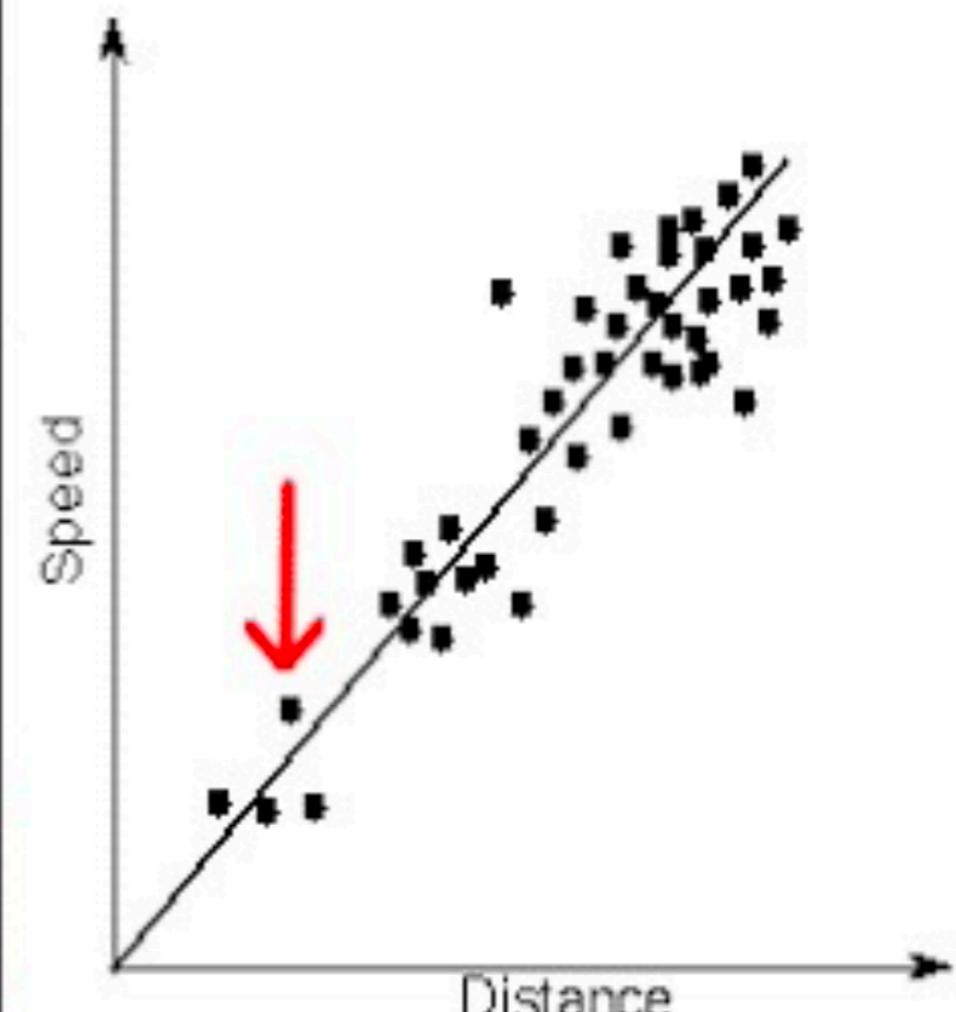
Mt. Wilson, CA 1927

Andromeda is far away

Edwin
Hubble,
Caltech



Hubble Law
recession speed = $H_0 \times$ distance

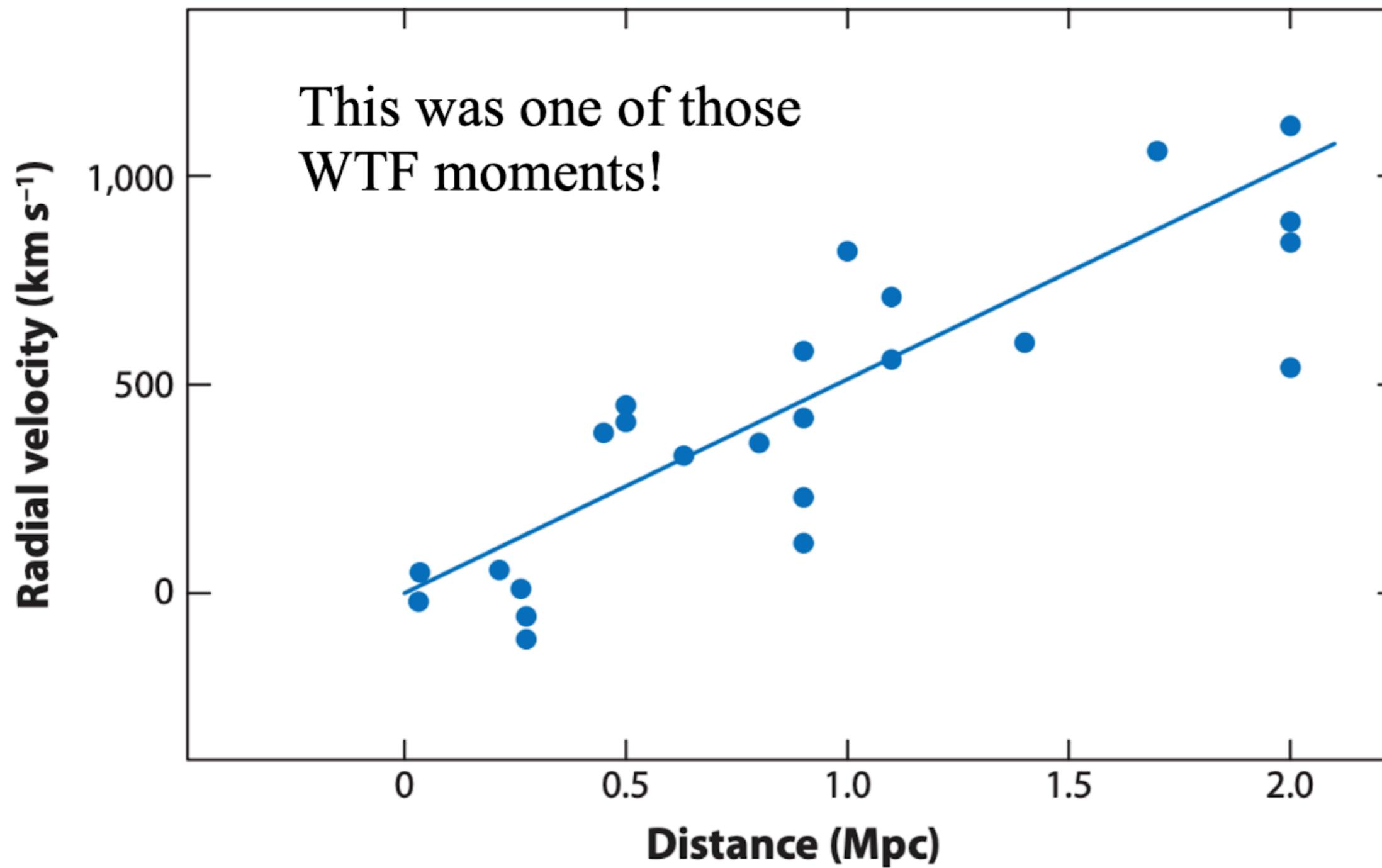


With Cepheid tech
Hubble showed Andromeda was far away

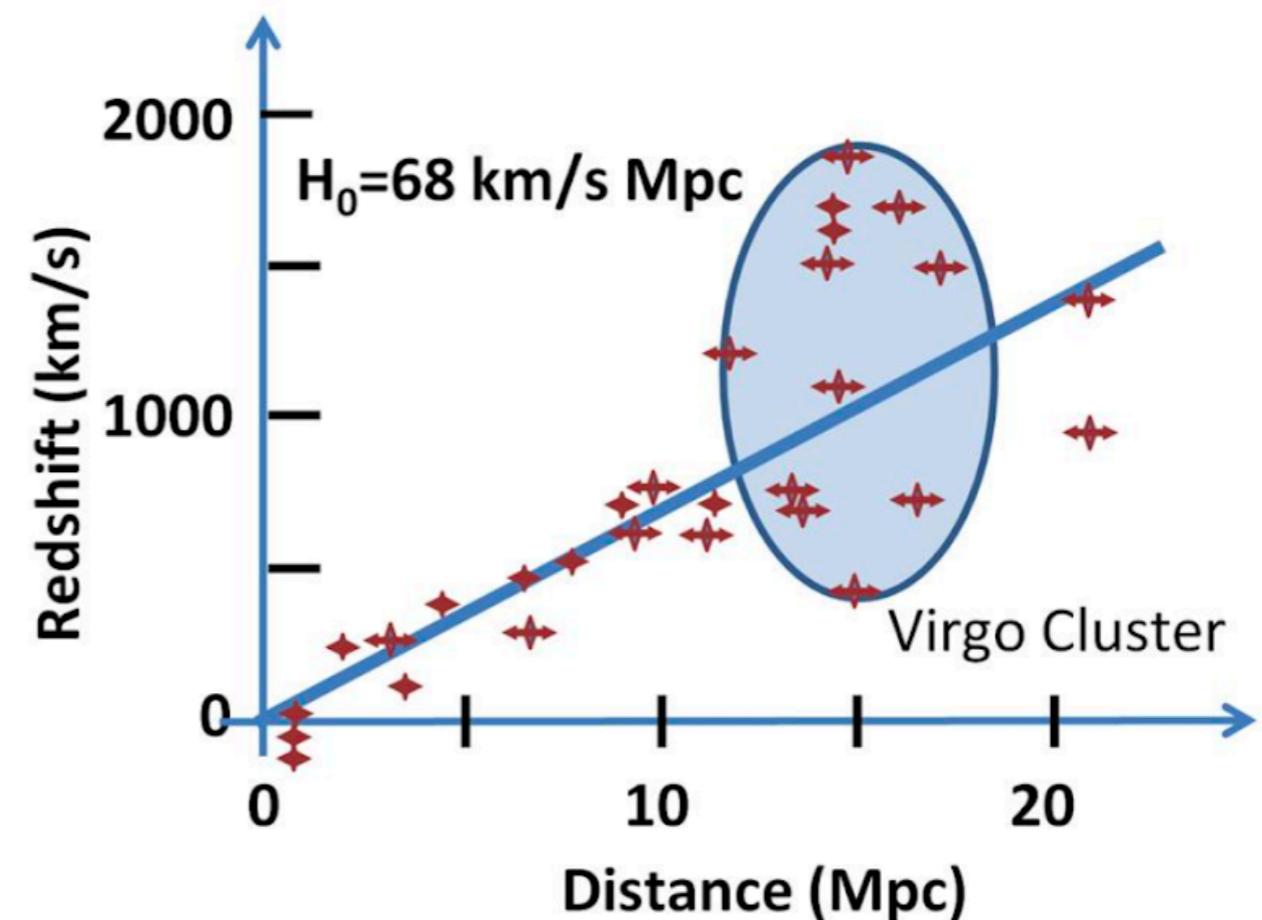
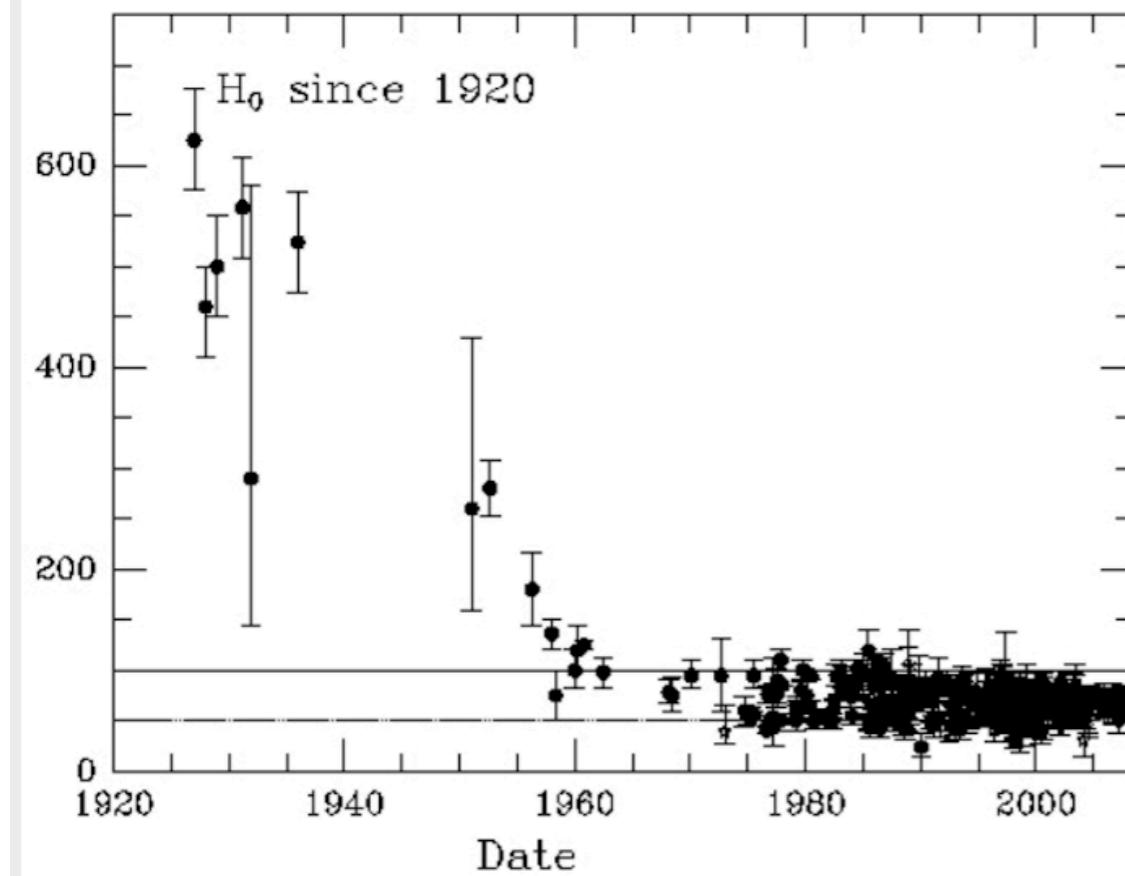
Also

Hubble's Expansion

All of the nearby galaxies are expanding



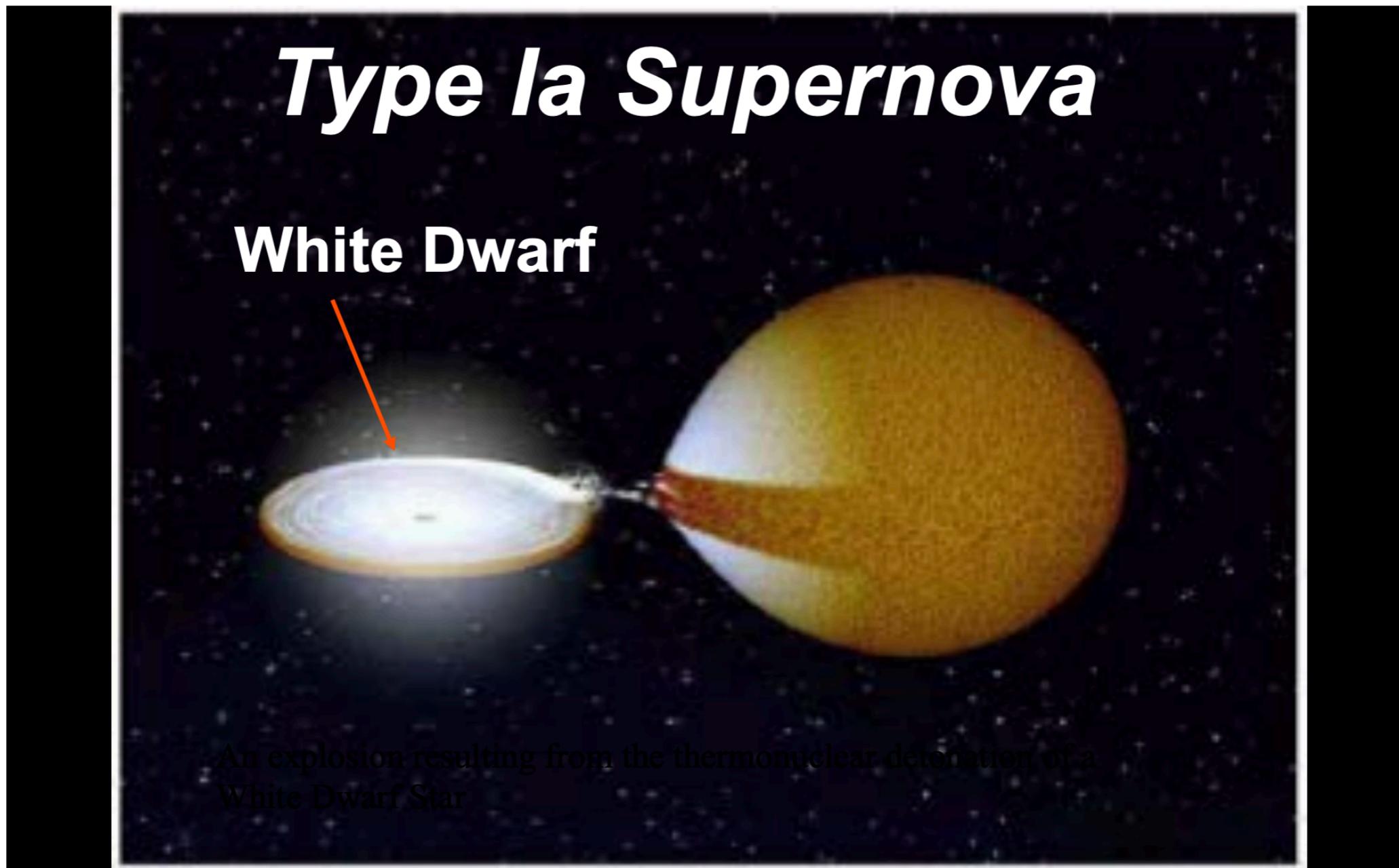
Controversy of Hubble's Measurement



Hubble was waaaay off

Factoring in galaxy clusters
and other effects are needed

Nowadays

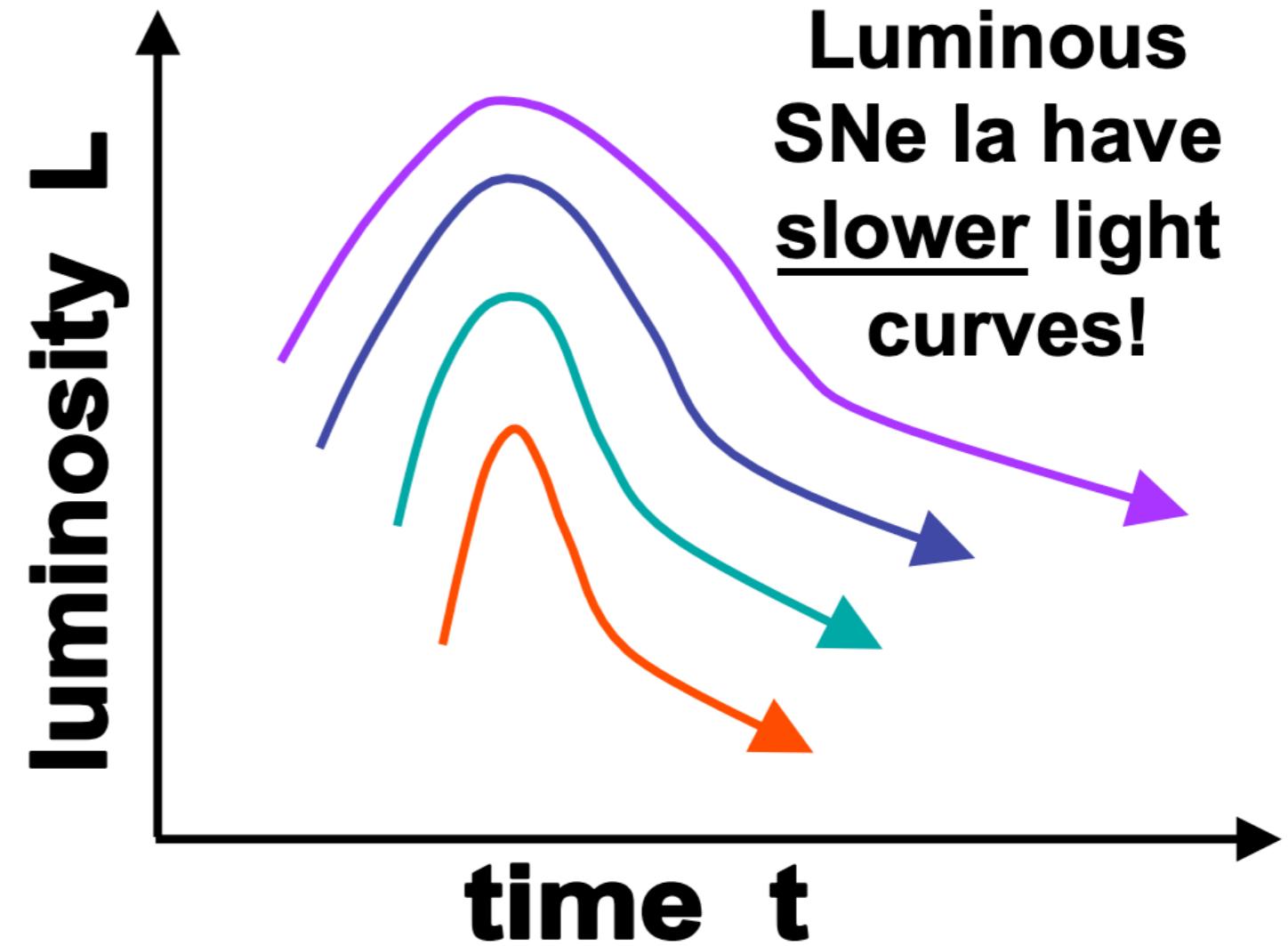


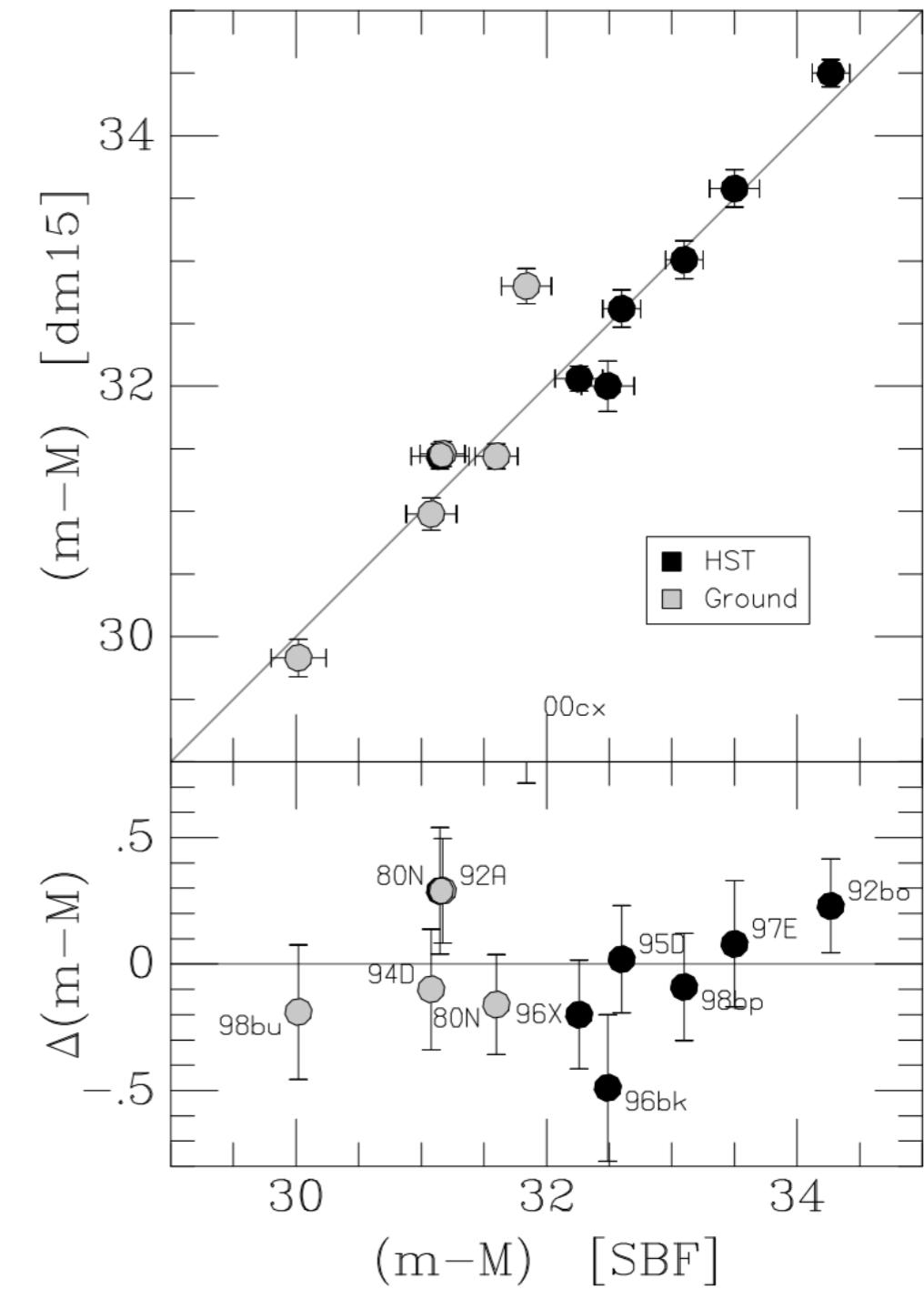
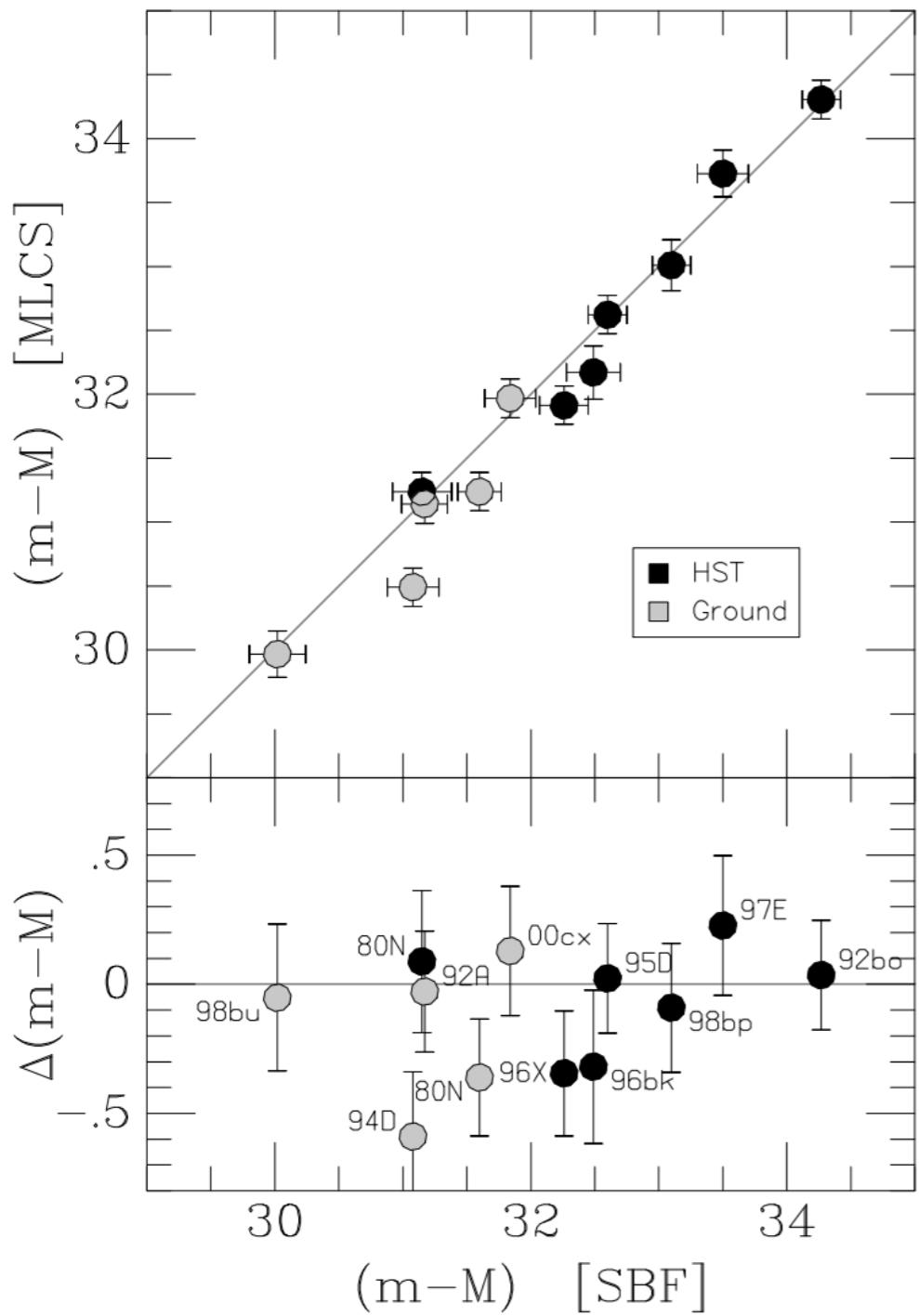
An explosion resulting from the thermonuclear runaway of a white dwarf near $M(\text{Chandrasekhar})$

Calibrating the Nearly Standard Candle

- Phillips (1993), Riess + (1995), Hamuy+ (1995): established L vs. light-curve shape correlation with ~ 10 nearby SNe Ia
- Use it to standardize other SNe Ia
- Measured colors give reddening and extinction
- Accurately calibrate individual SNe!

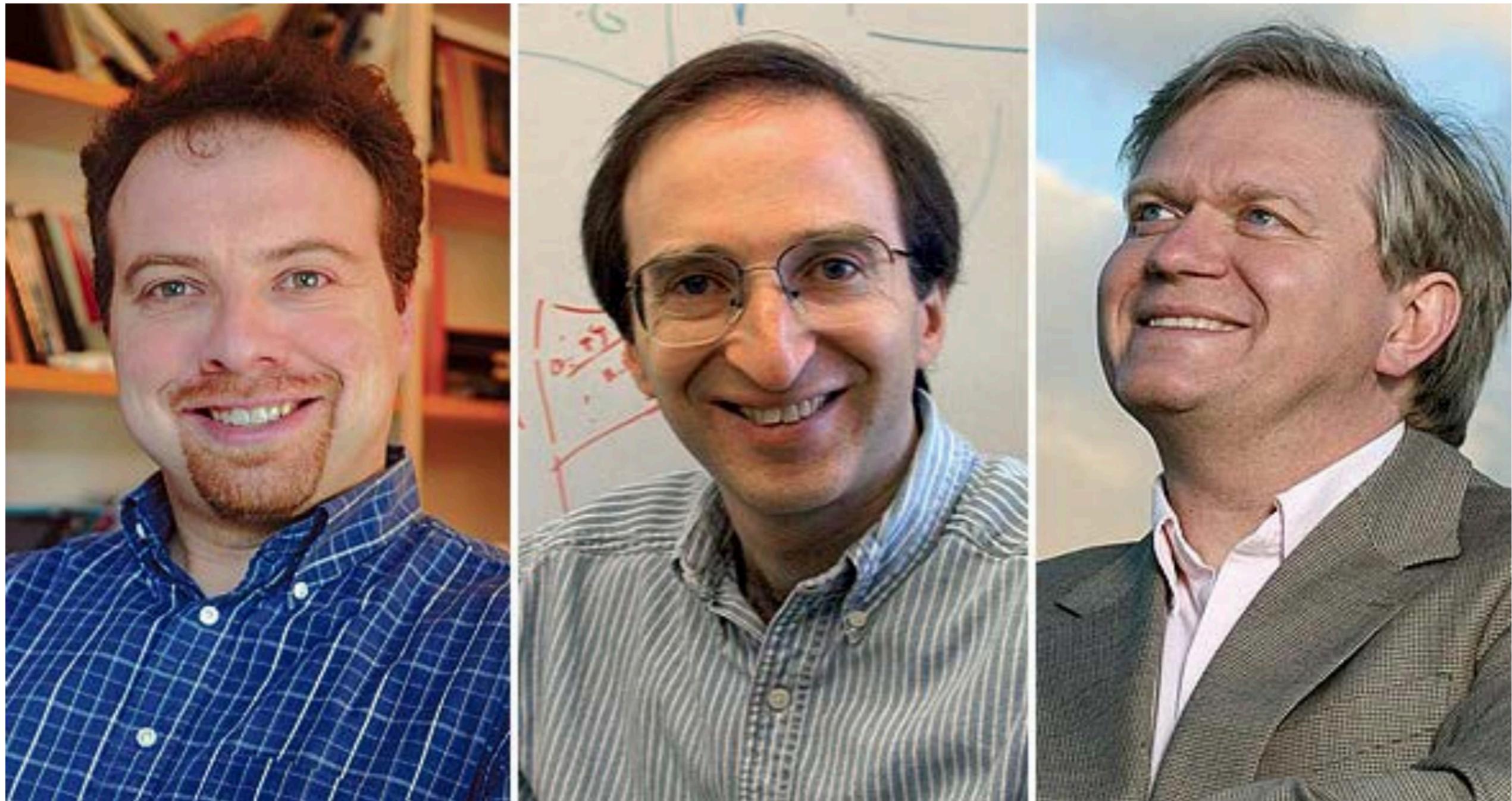
Absolute light curves of SN Ia in galaxies of known distance





Calibration

Studies of Universe's Expansion Win Physics Nobel



Johns Hopkins University; University Of California At Berkeley; Australian National University

From left, Adam Riess, Saul Perlmutter and Brian Schmidt shared the Nobel Prize in physics awarded Tuesday.

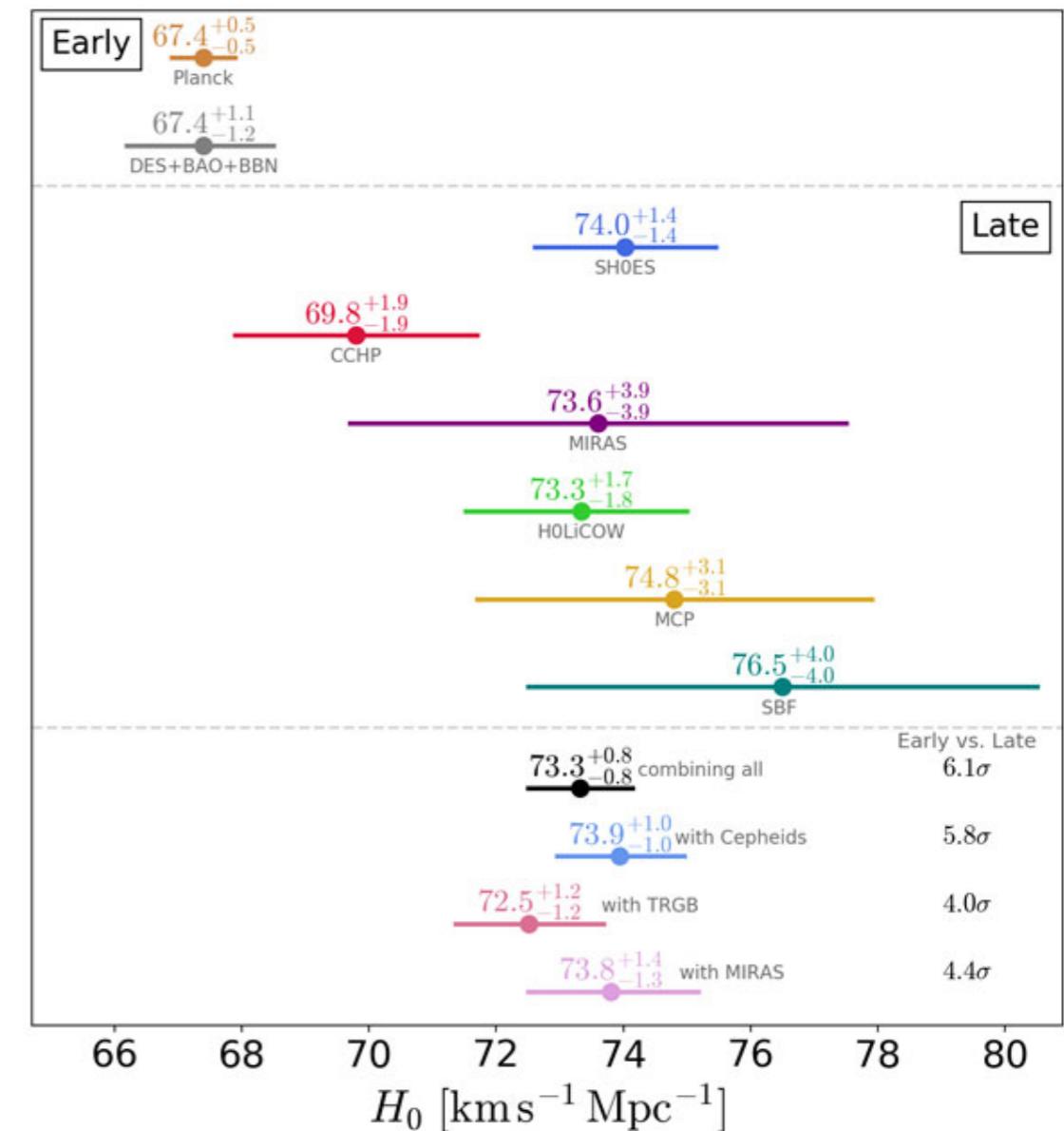
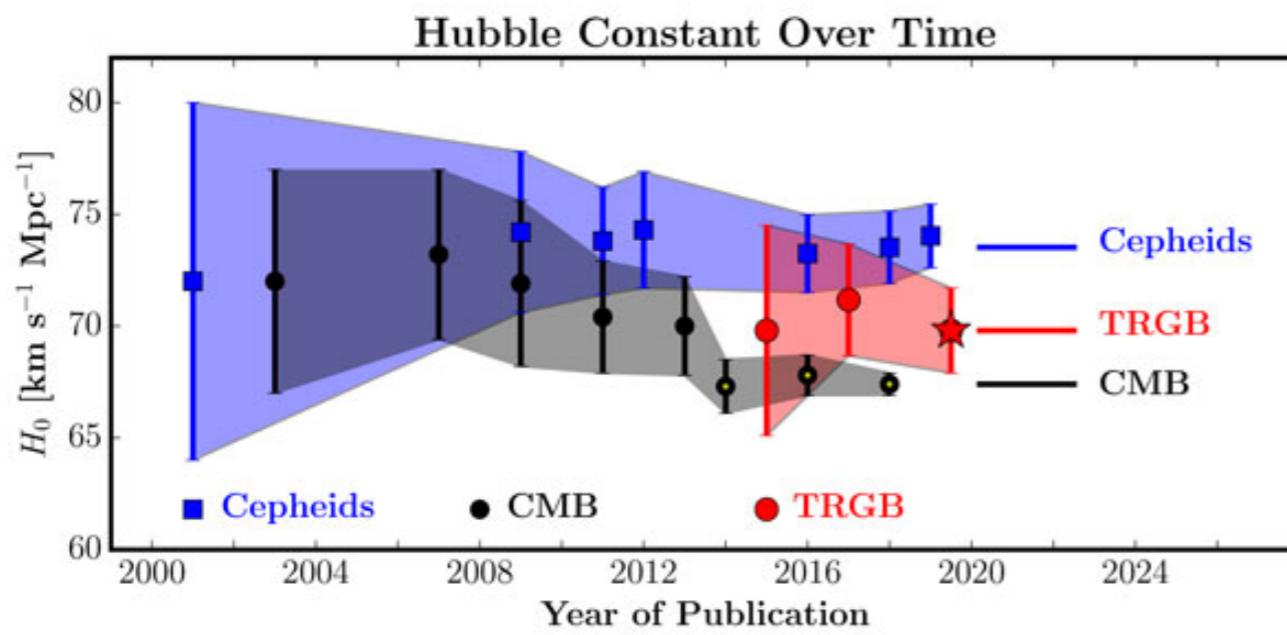
By [DENNIS OVERBYE](#)

Published: October 4, 2011

2011 Nobel Prize in Physics

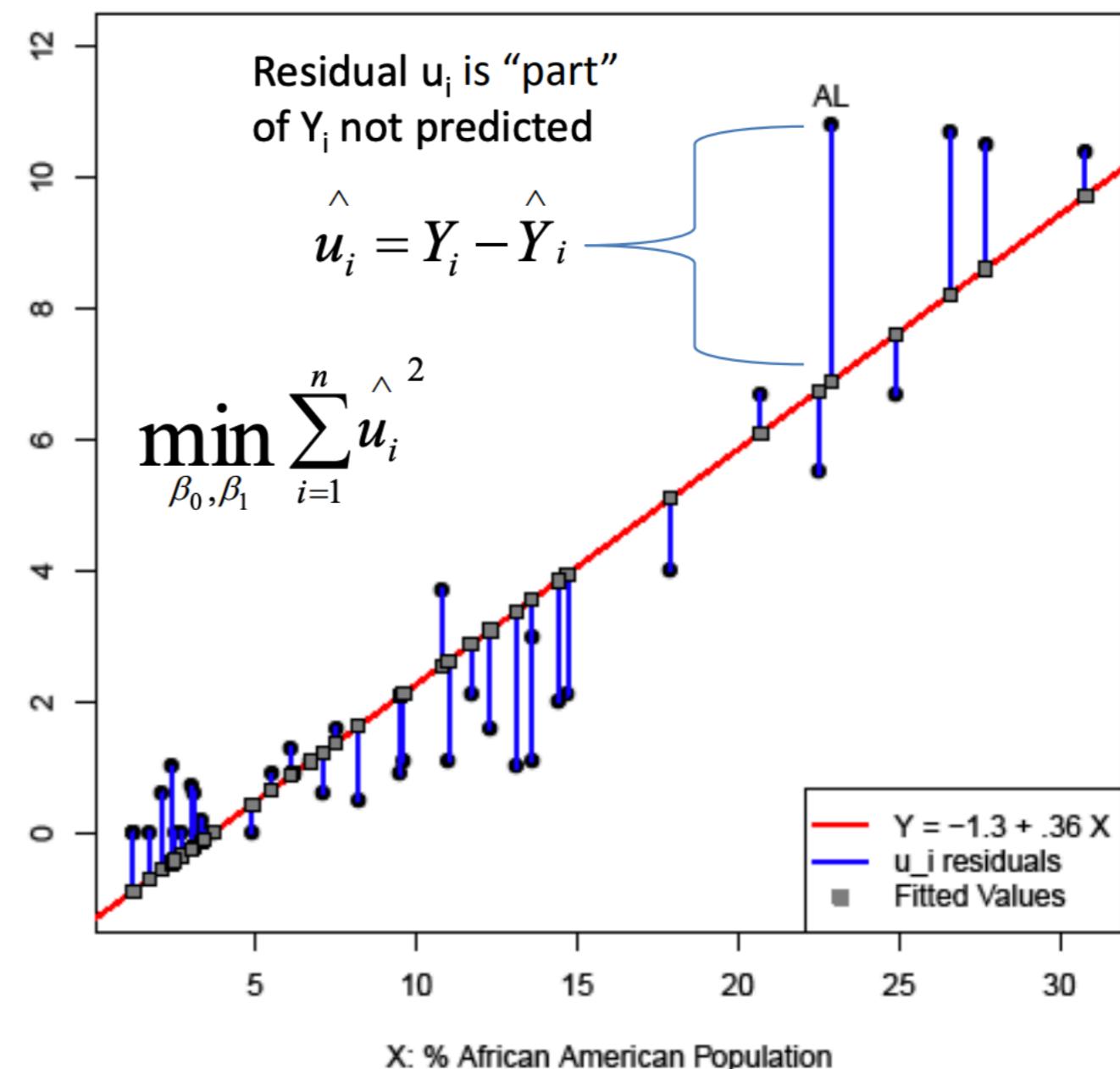
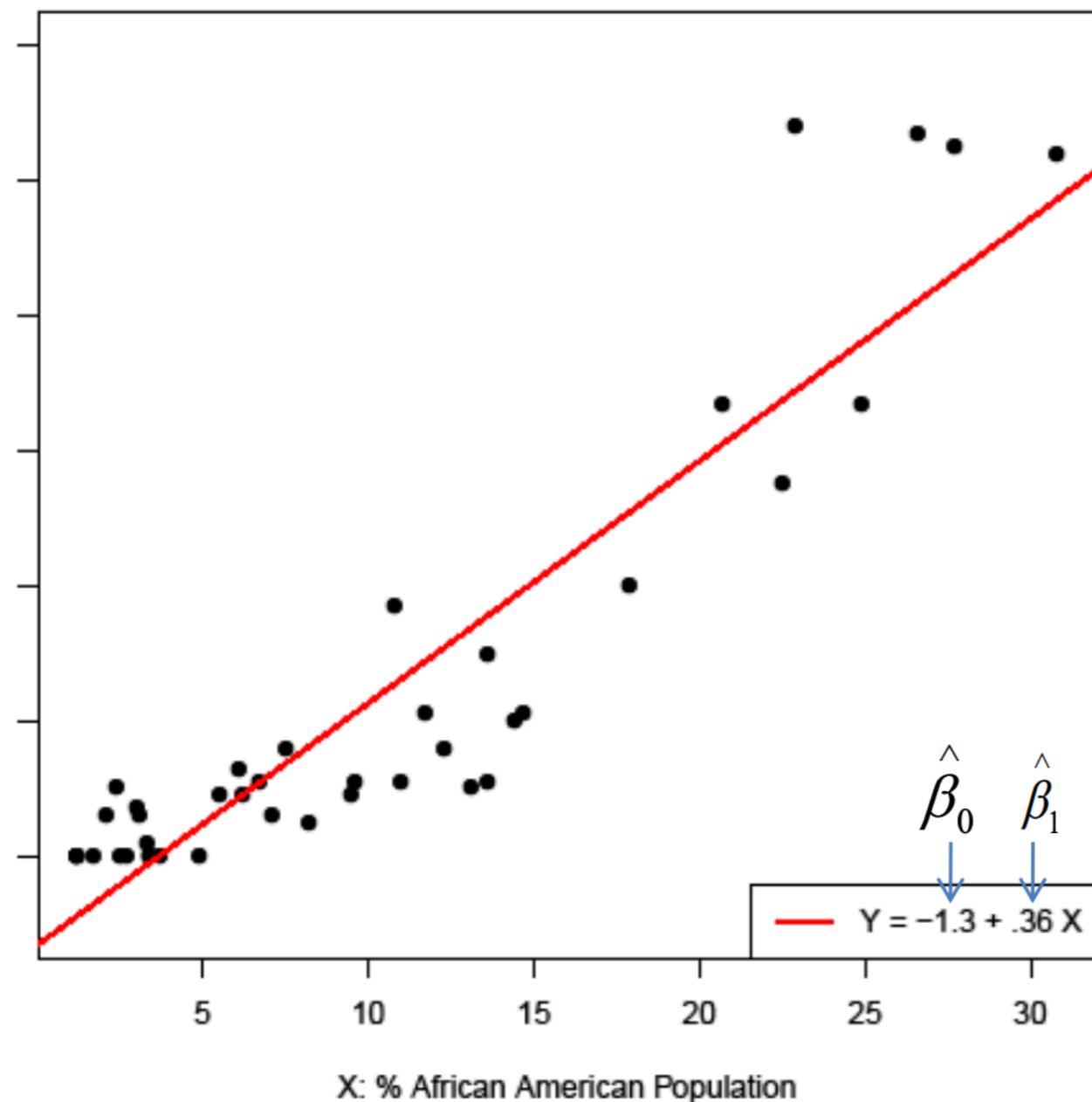
However there appears to be tension¹⁴

- Current Supernovae+Cepheid measurements
 - Indicate a larger value of Hubble's constant
- CMB indicates a smaller value
- Lets do the fit ourselves!

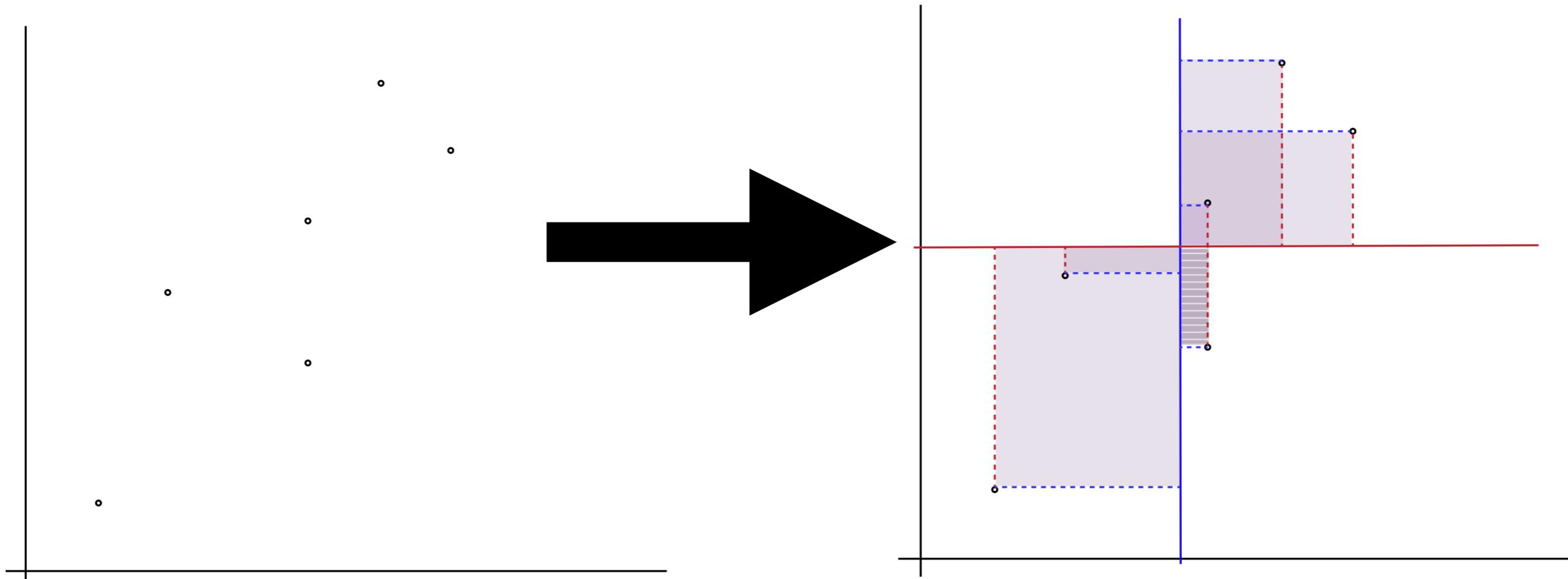


Linear Regression from scratch

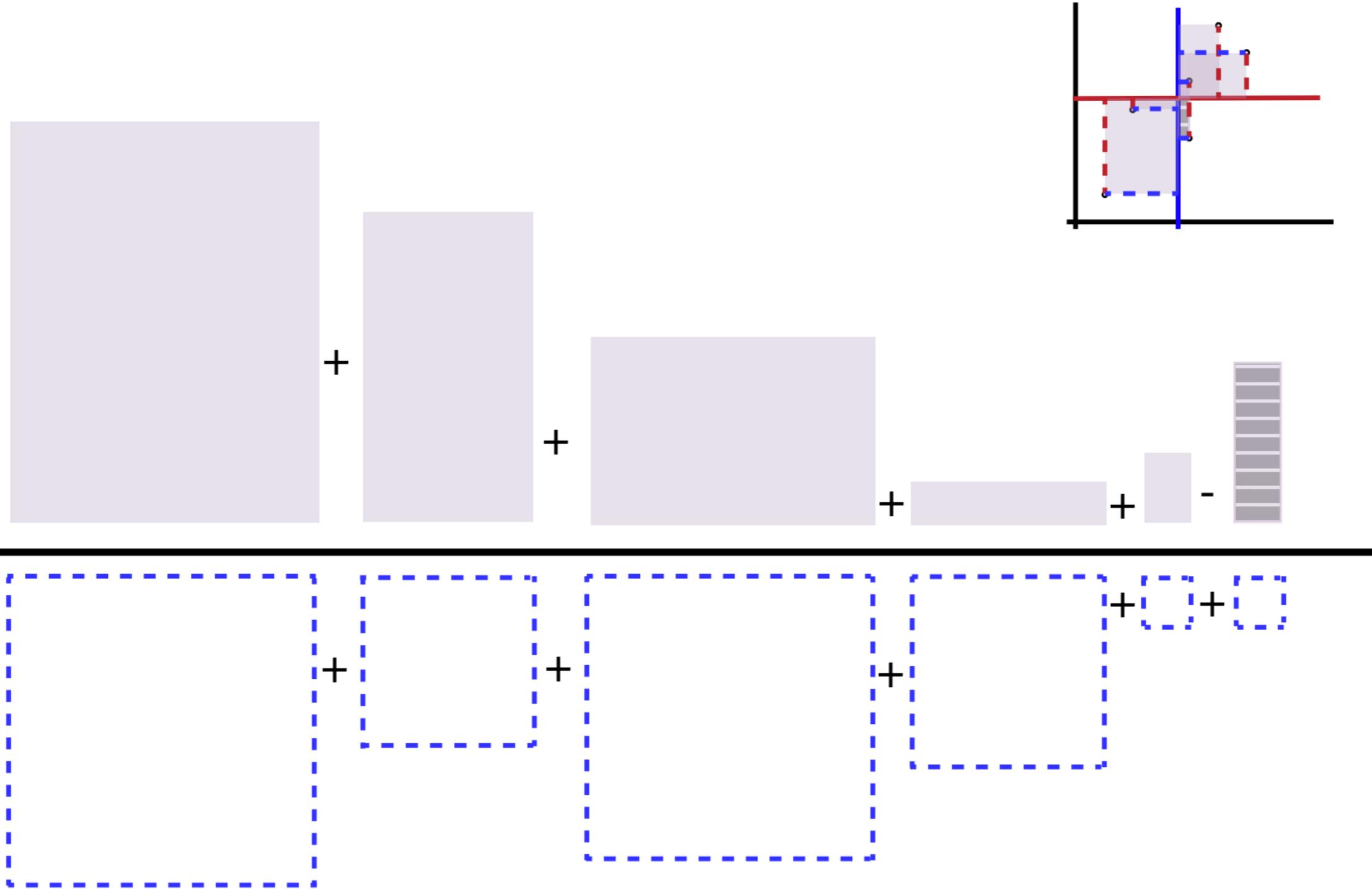
What do we want to¹⁶ minimize?



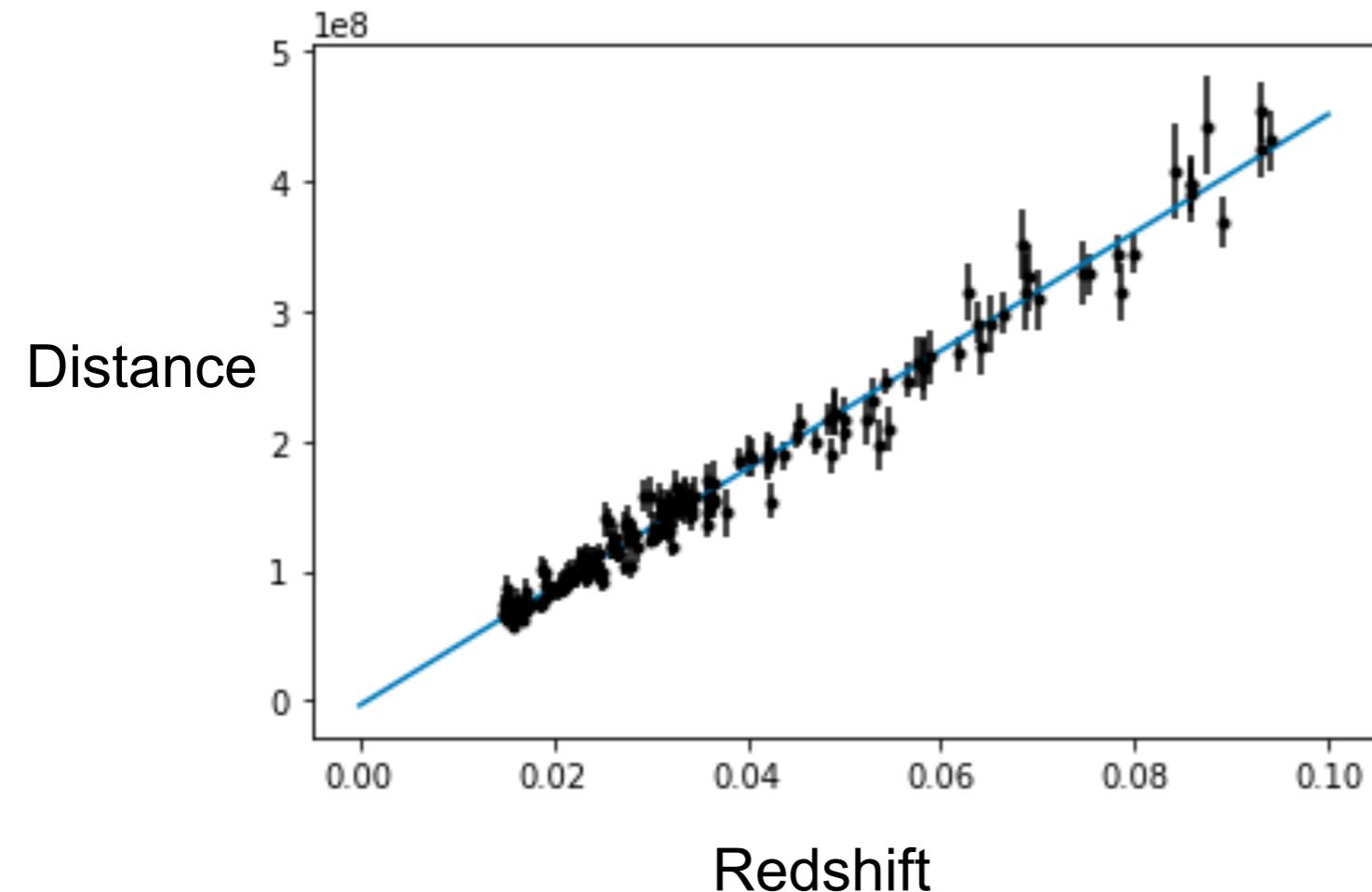
Visualizing Derivation



Visualizing Derivation



Hubble's Measurement



- $A = \text{Distance (pc)} / \text{redshift} = (\text{Mpc}/10^6) / (c/v) / \text{Redshift}$
- Redshift = $v/c \Rightarrow \text{Redshift} = v / c(3 \times 10^6 \text{ km/s}) = (v \text{ in km/s}) / 3 \times 10^6 \text{ km/s}$
- $A = 1/10^6 \text{ (Distance in Mpc)} / (v \text{ in Km/s} / 3 \times 10^6 \text{ km/s}) = 1/(3 \times 10^6 \times 10^6) \text{ (Distance in Mpc)} / (\text{velocity in km/s})$
- Hubble's Constant: $H_0 = 67 \text{ km/s/Mpc}$ which is velocity in km/s over distance in Mpc
- $H_0 = (3 \times 10^6 \times 10^6)/A$ With error propagation $\rightarrow \sigma_h = (3 \times 10^{12}/A^2)\sigma_A$

Regression Forms

$$A = \frac{Cov(X, Y)}{Var(X)} \text{ or } \frac{\sum xy}{\sum x^2} \text{ or } \frac{\sum_{i=1}^n (X_i Y_i) - \frac{\sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{n}}{\left(\sum_{i=1}^n X_i\right)^2} \text{ or } \frac{\sum_{i=1}^n (X_i Y_i) - n \bar{X} \bar{Y}}{\sum_{i=1}^n (X_i^2) - n \bar{X}^2} \text{ or } \frac{\frac{1}{n} \sum_{i=1}^n (X_i Y_i) - \bar{X} \bar{Y}}{\frac{1}{n} \sum_{i=1}^n (X_i^2) - \bar{X}^2} \text{ or } \frac{\overline{(XY)} - \bar{X} \bar{Y}}{\overline{(X^2)} - \bar{X}^2}$$

- There are a lot of different ways to write the regression

More Regression Forms

$$\sum x^2 = SS_x = (n-1)Var(X) = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2) - n\bar{X}^2$$

$$\sum y^2 = SS_y = (n-1)Var(Y) = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (Y_i^2) - n\bar{Y}^2$$

$$\sum xy = S_{xy} = (n-1)Cov(X, Y) = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n (X_i Y_i) - n\bar{X}\bar{Y}$$

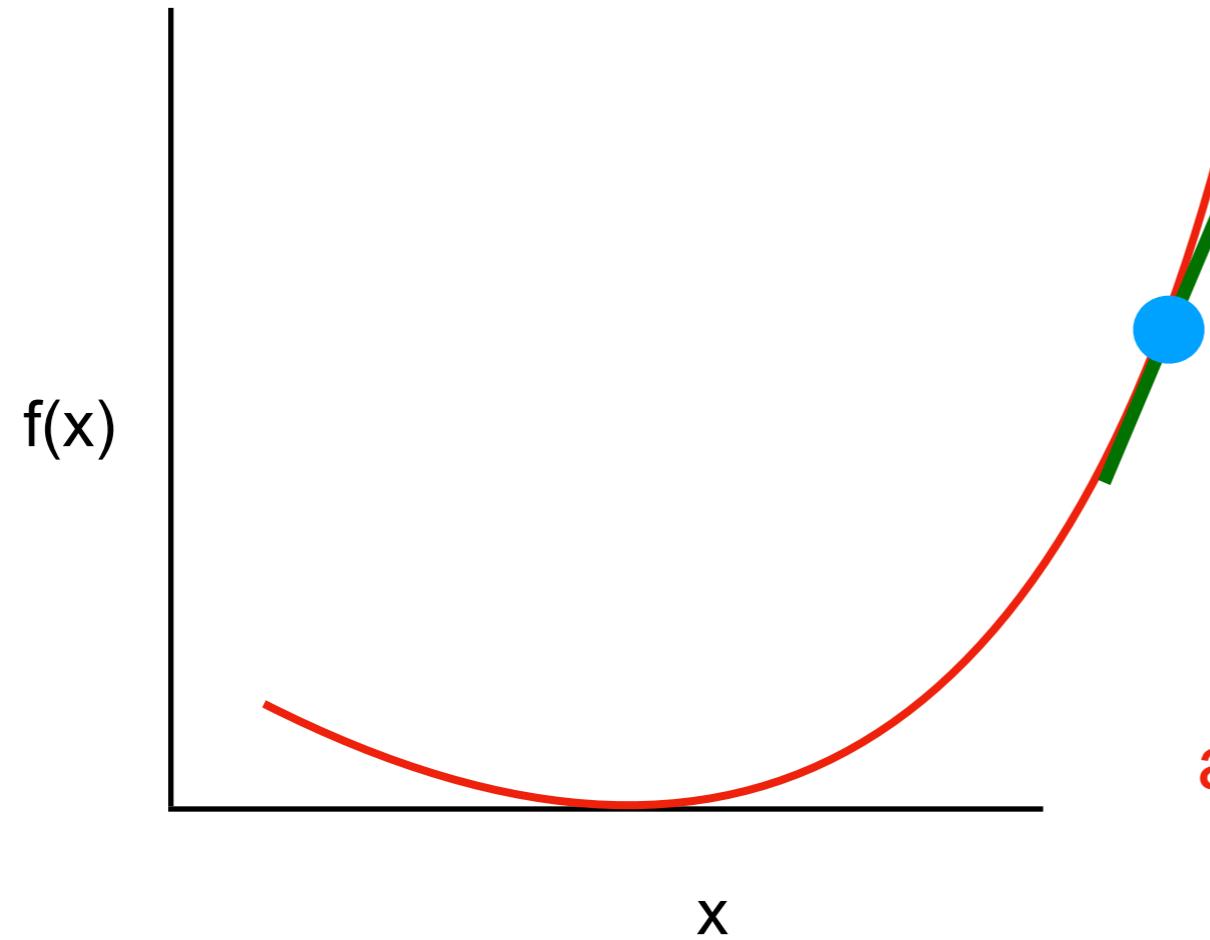
- Here we refer to SS as the sum of the squares

Gradient Descent



- The ideas are very similar to real life
- How do we get to the minimum as fast as possible

Gradient



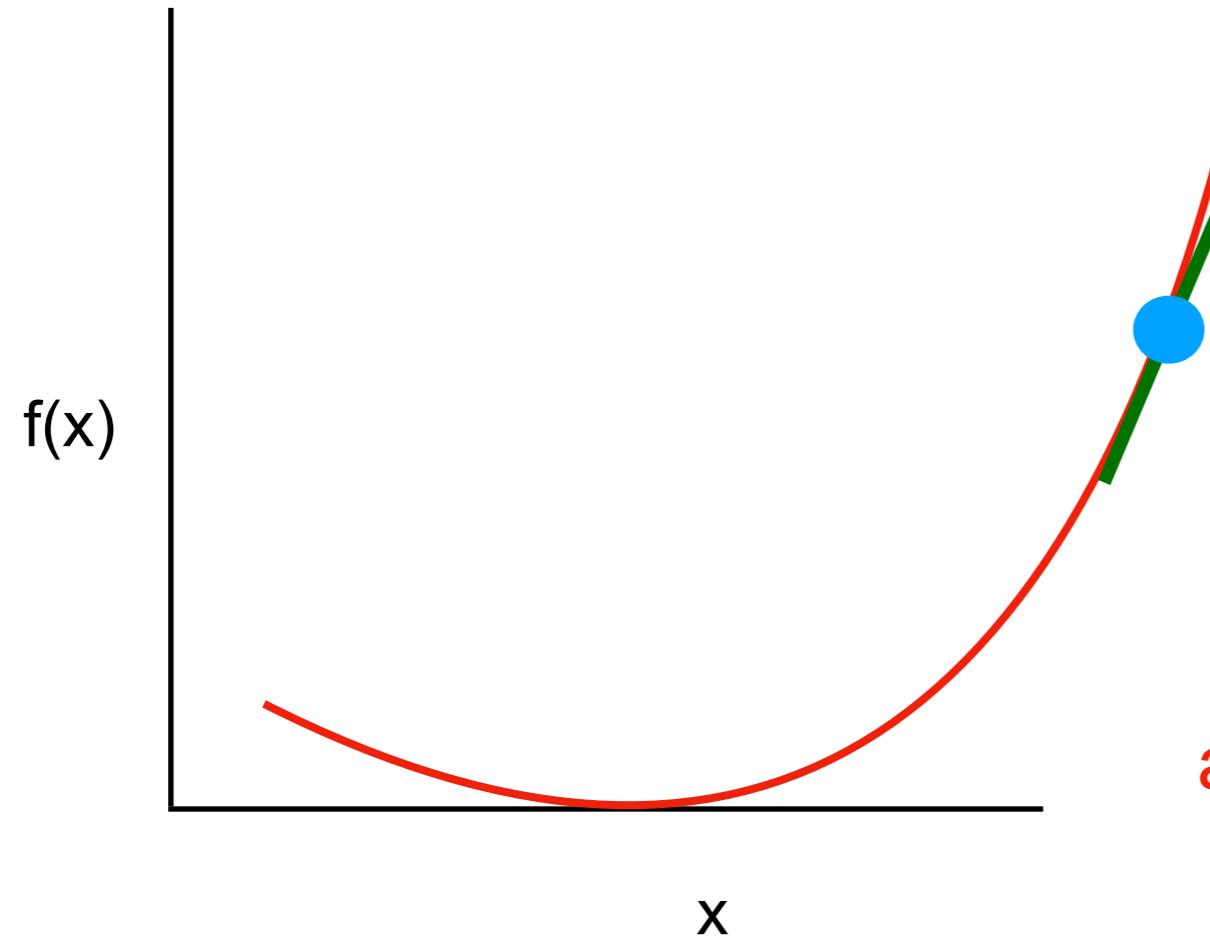
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



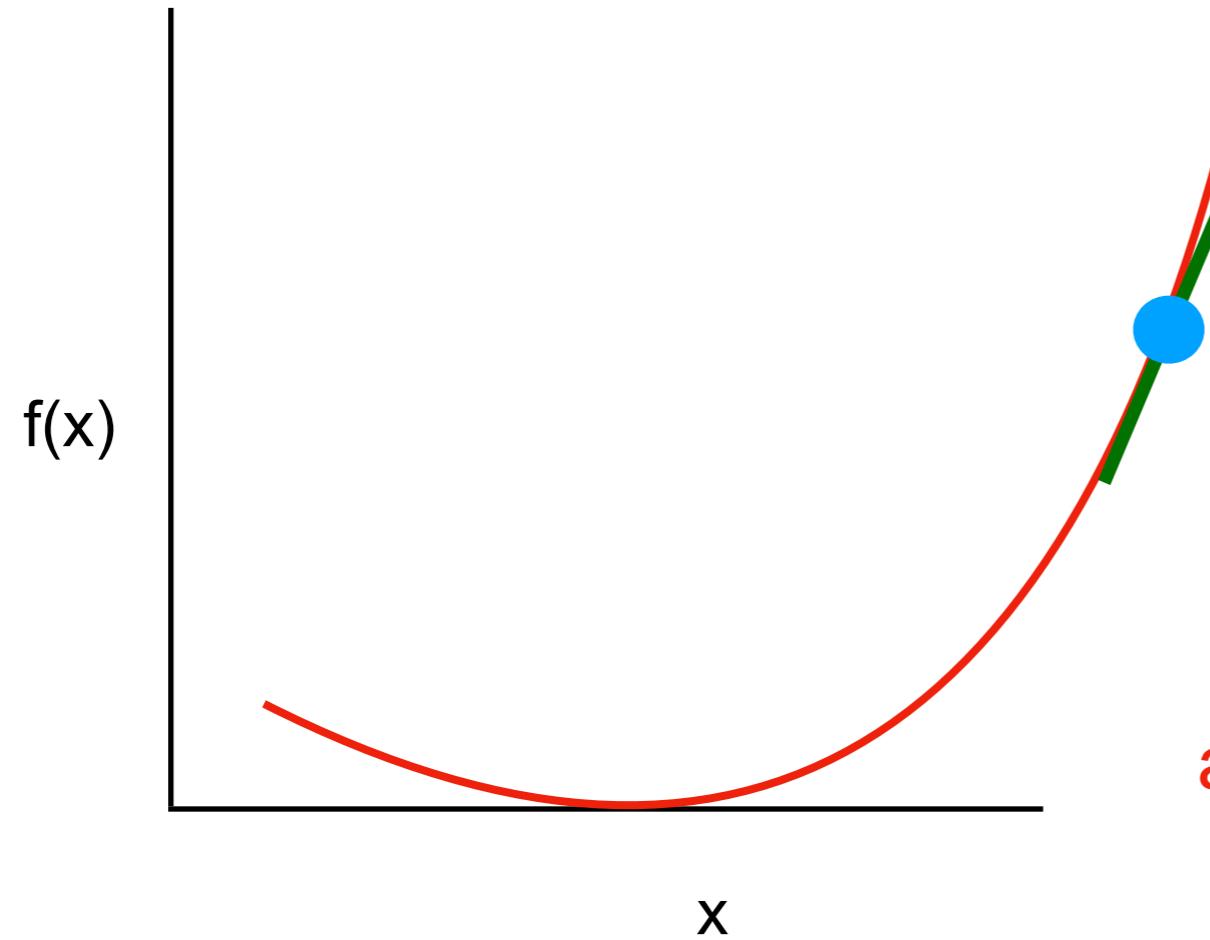
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



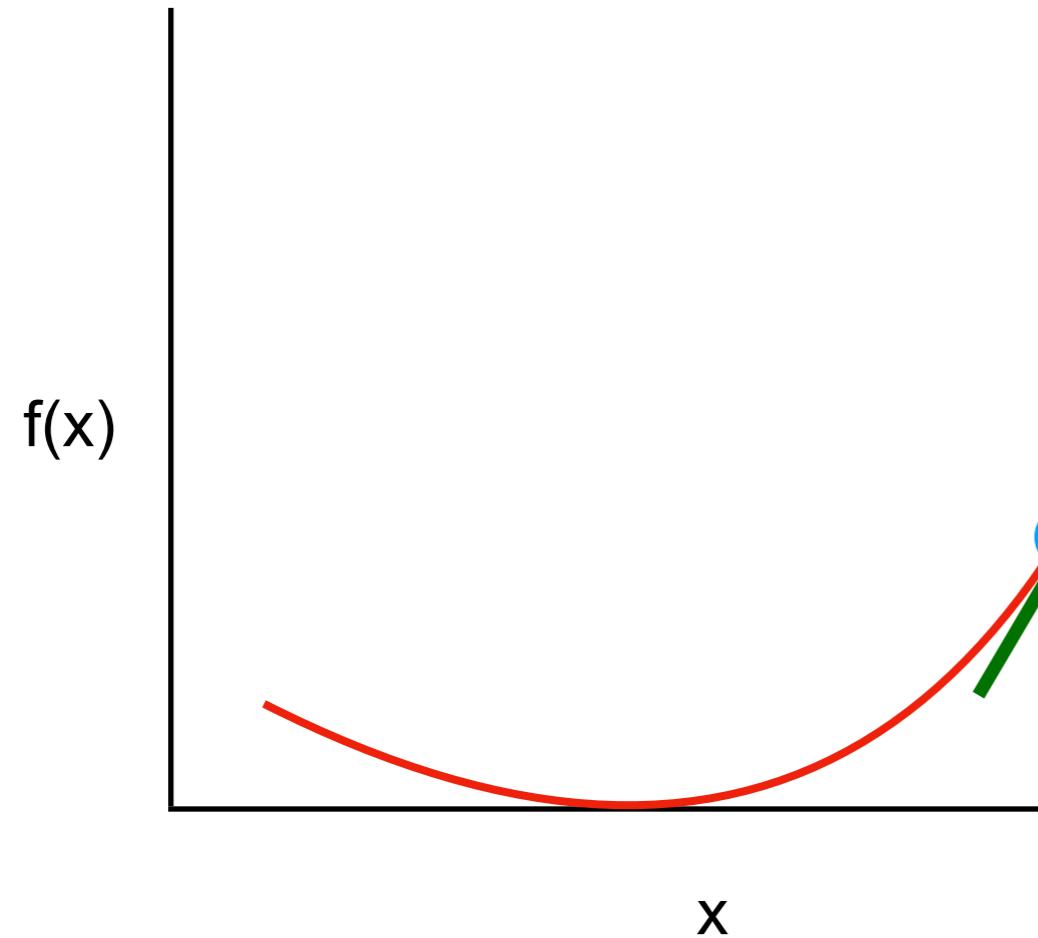
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



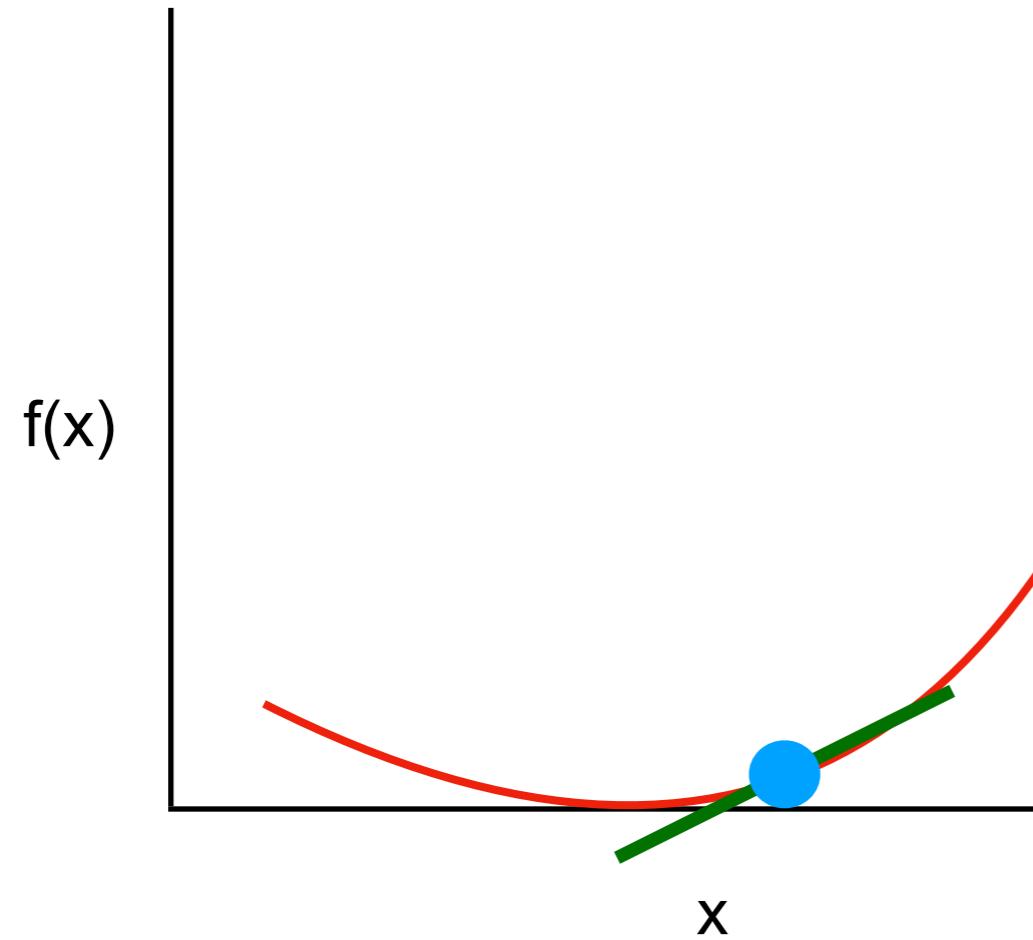
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



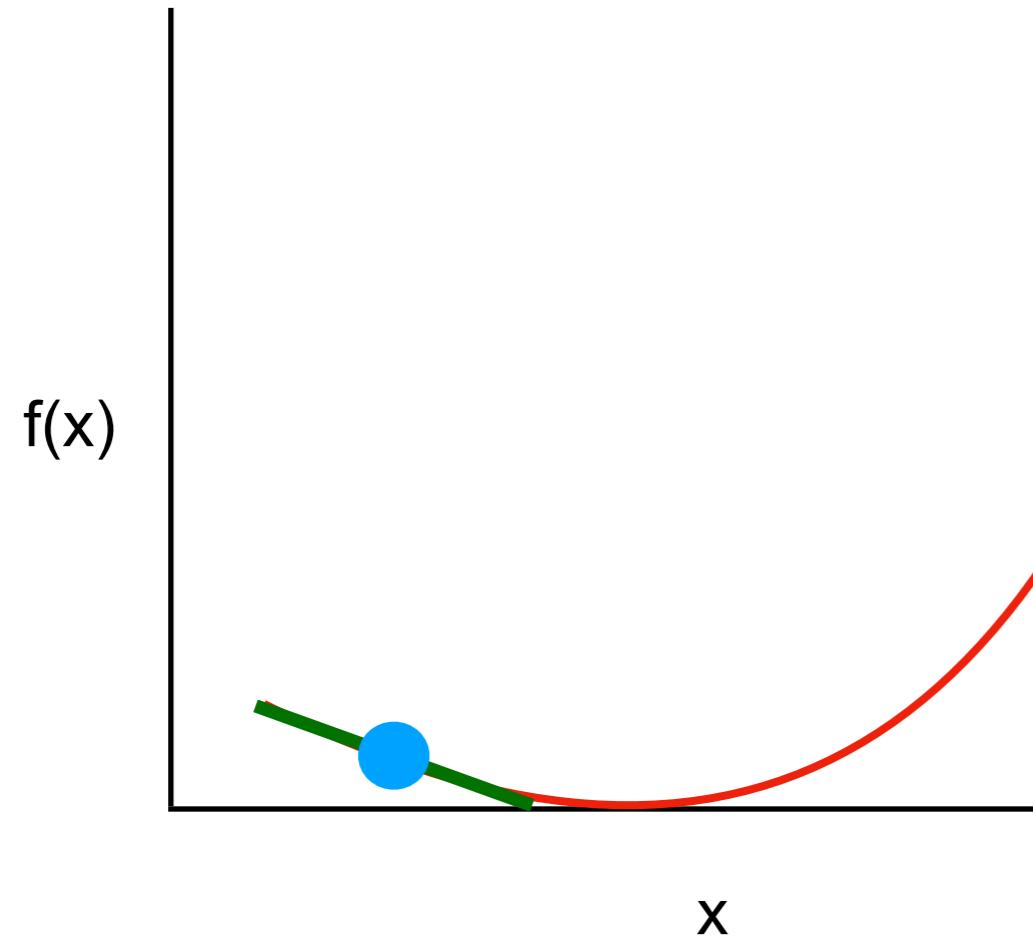
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



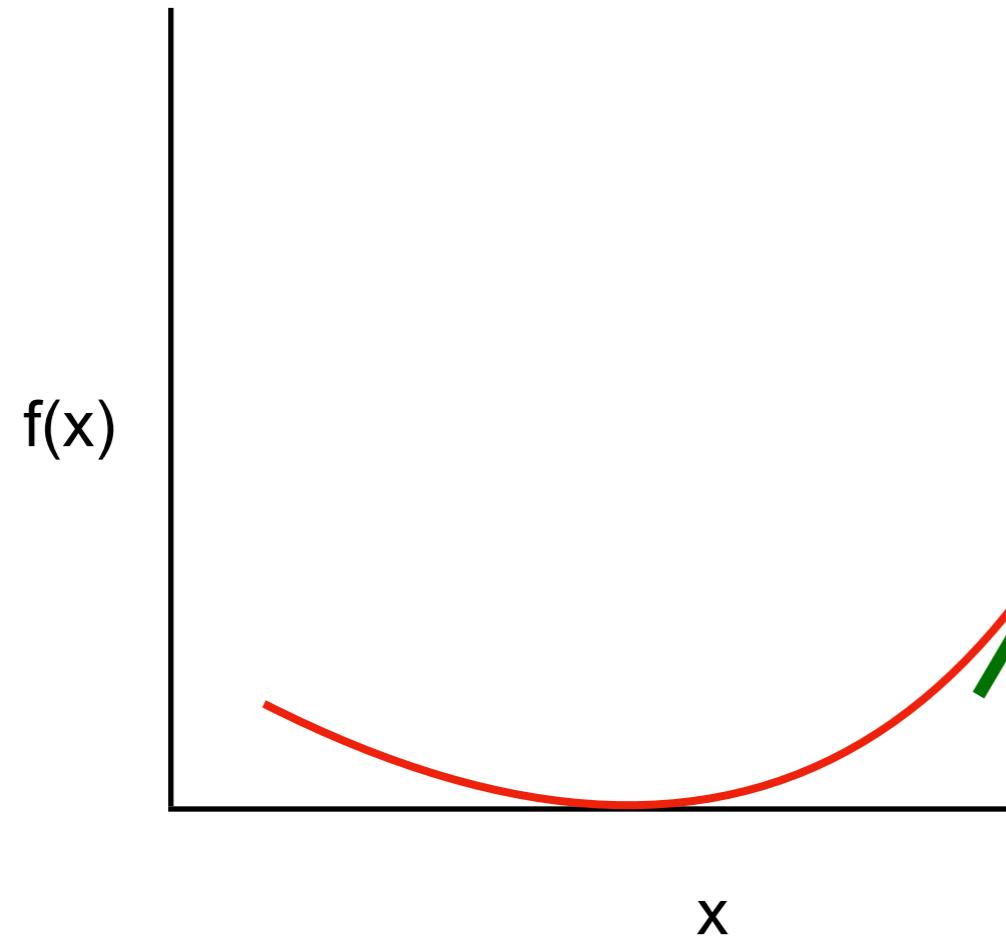
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U$$

Gradient



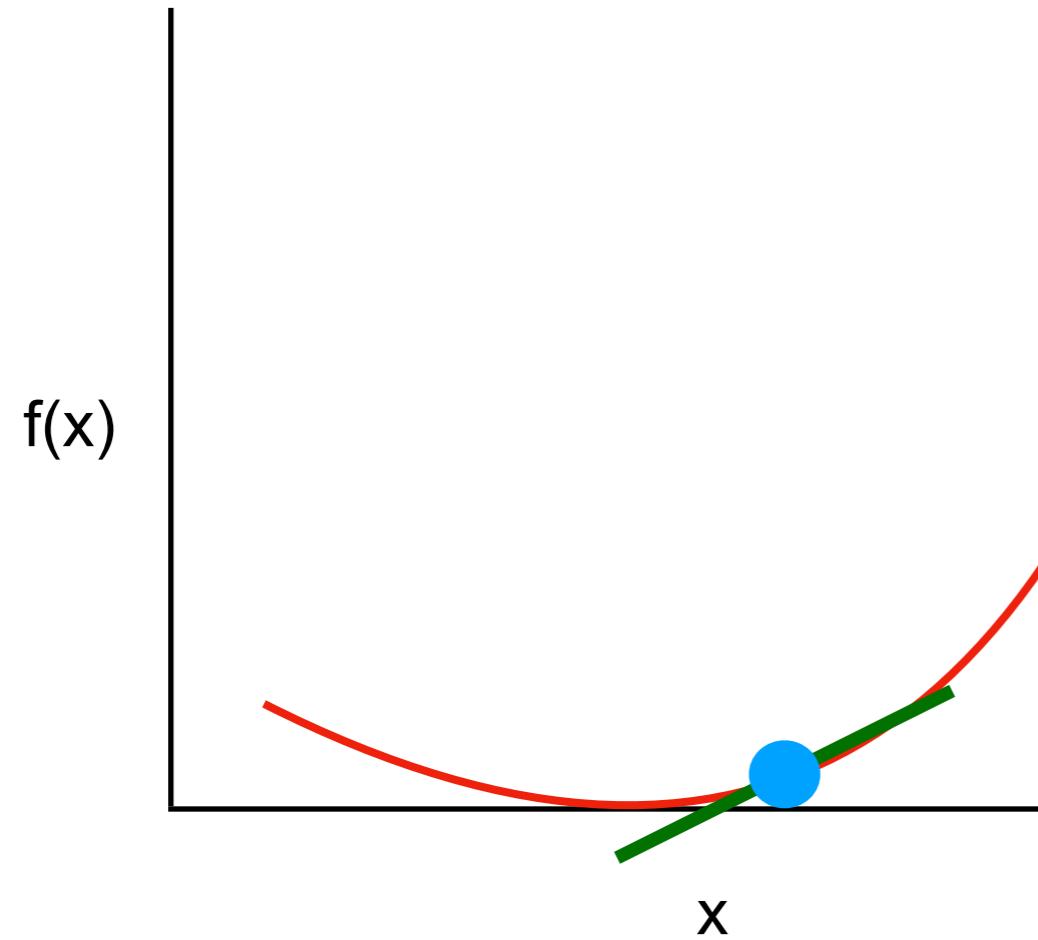
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U + \text{Dampen}$$

Gradient



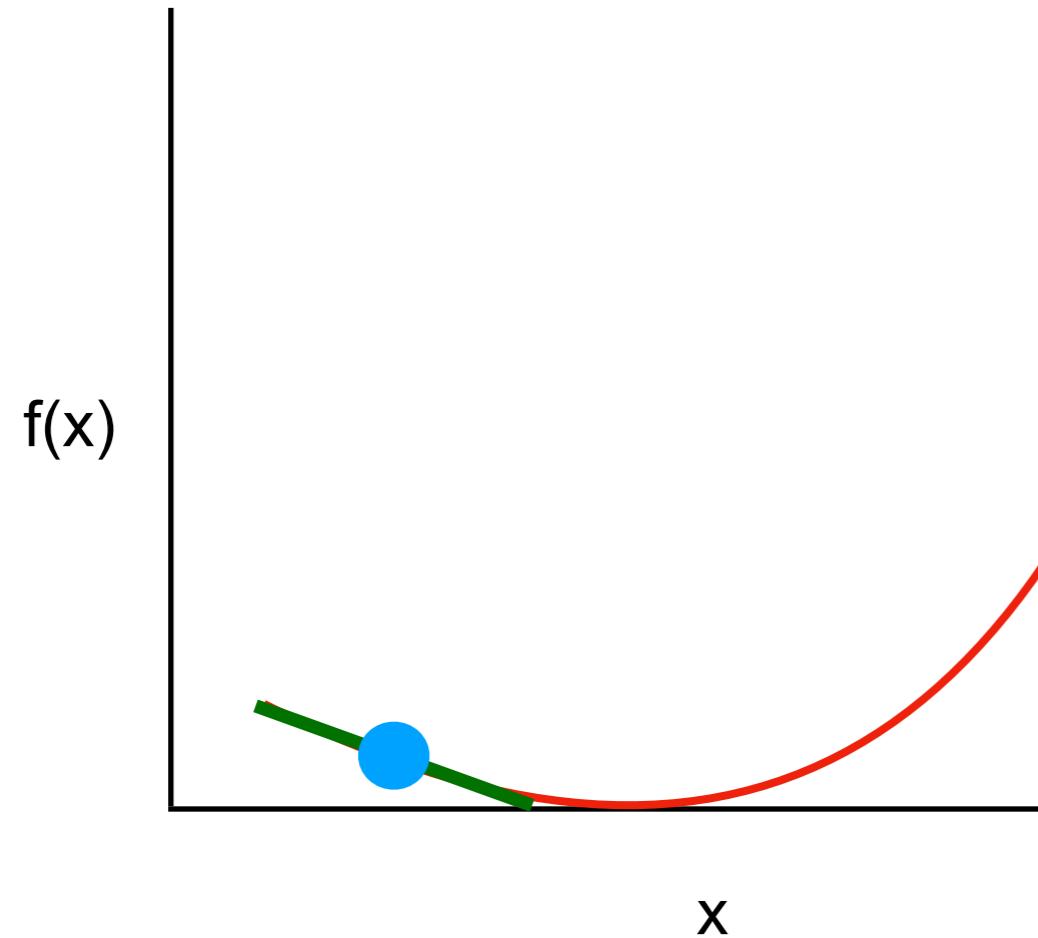
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U + \text{Dampen}$$

Gradient



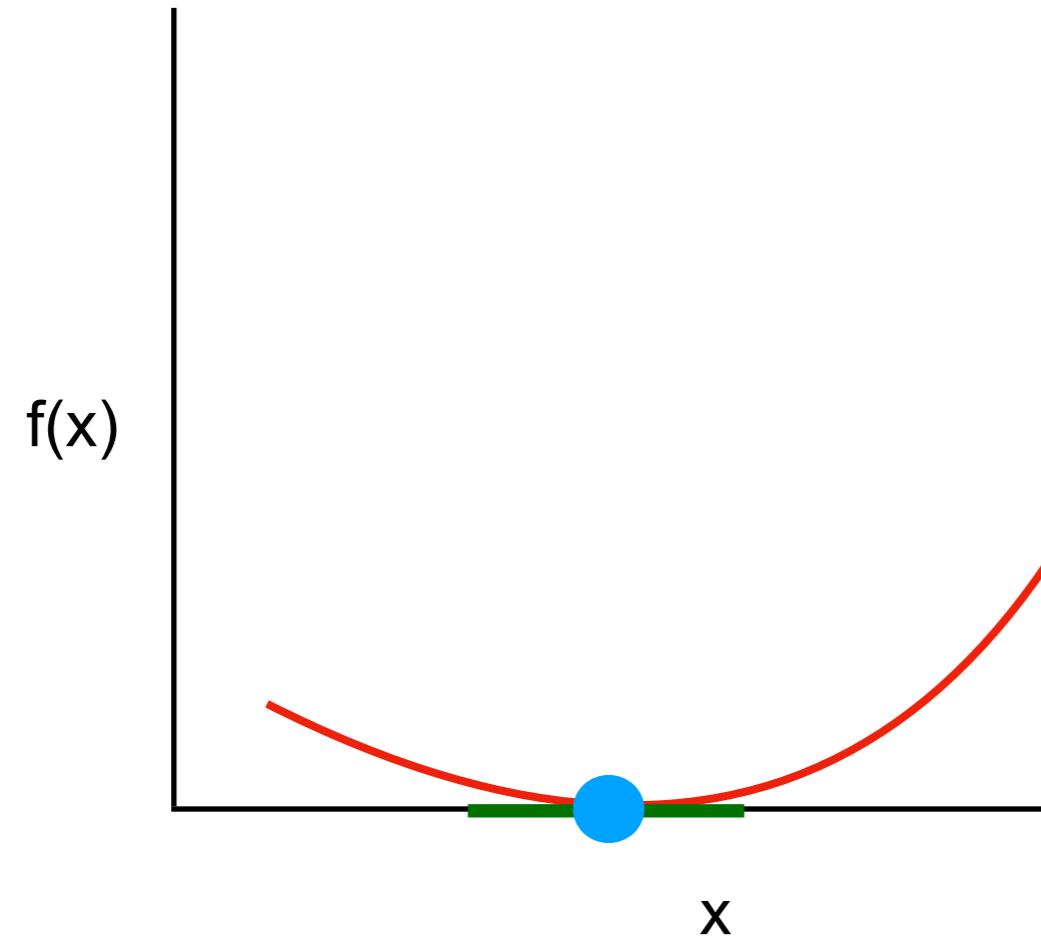
Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U + \text{Dampen}$$

Gradient



Simplest approach let
a ball roll down a potential

$$\mathbf{F} = -\nabla U$$

$$\vec{v} = \vec{v}_0 - \delta t \nabla U$$

$$\vec{x} = \vec{x}_0 + \delta t \vec{v} - \frac{1}{2} (\delta t)^2 \nabla U + \text{Dampen}$$

Newton's Step

$$f(x + h) = f(x) + h \frac{df}{dx}(x) + \frac{h^2}{2} \frac{d^2 f}{dx^2}$$

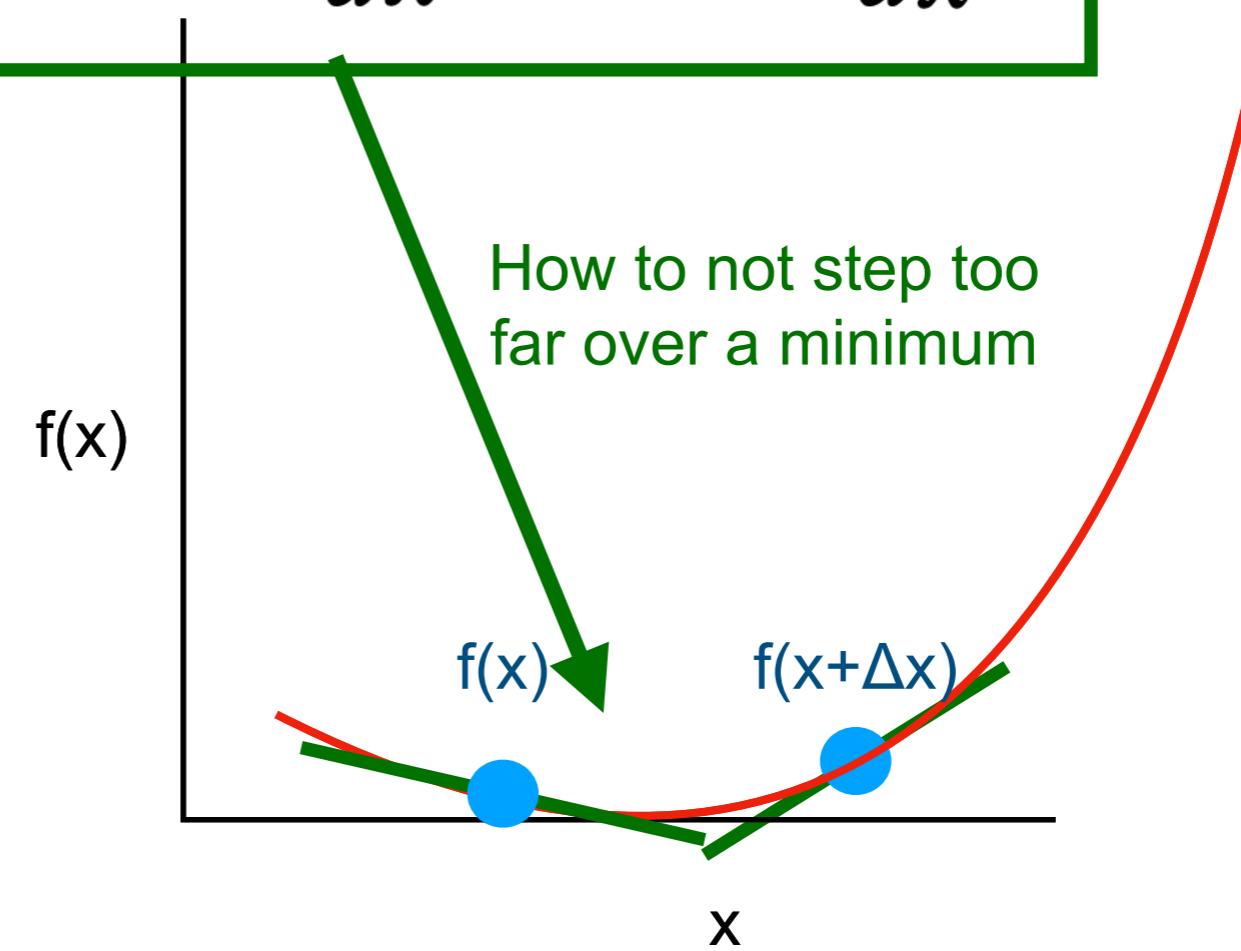
$$\frac{f(x + h) - f(x)}{h} = 0 = \frac{df}{dx}(x) + \frac{h}{2} \frac{d^2 f}{dx^2}$$

$$0 = \frac{df}{dx}(x) + \Delta x \frac{d^2 f}{dx^2}$$

$$\Delta x = -\frac{\frac{df}{dx}}{\frac{d^2 f}{dx^2}}(x)$$

$$\Delta x = -\frac{f'(x)}{f''(x)}$$

Newton Step

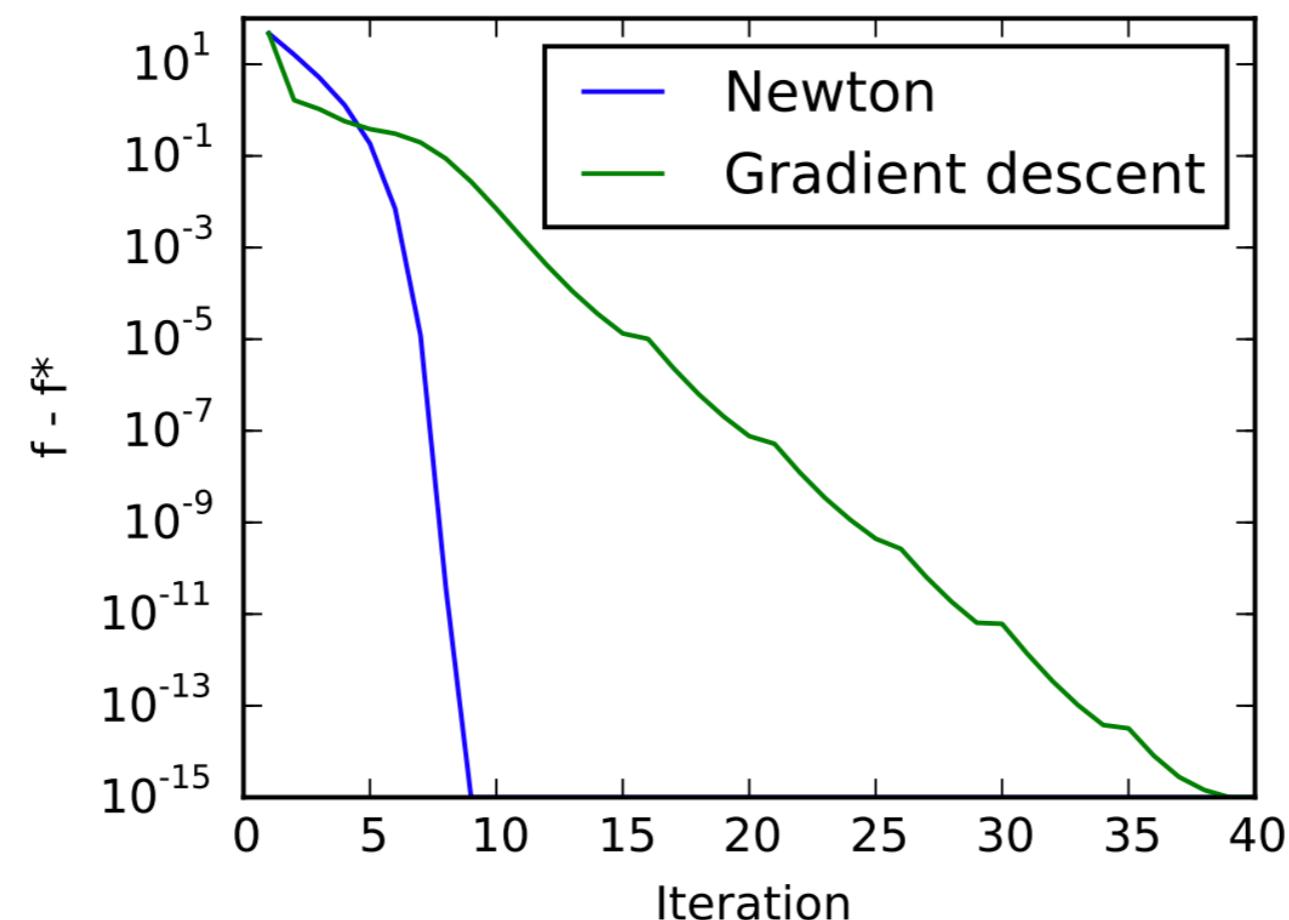


Optimal Minimizations

Example: Newton's method

$$f(x_1, x_2) = \exp(x_1 + 3x_2 - 0.1) + \exp(x_1 - 3x_2 - 0.1) + \exp(-x_1 - 0.1)$$

Convergence of Newton's method vs. gradient descent



Newton step

$$\Delta x_{\text{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

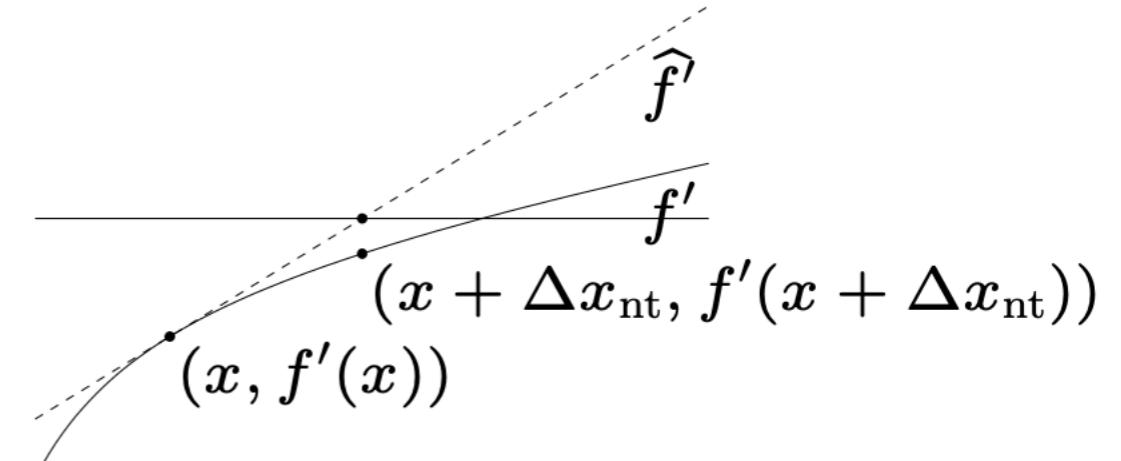
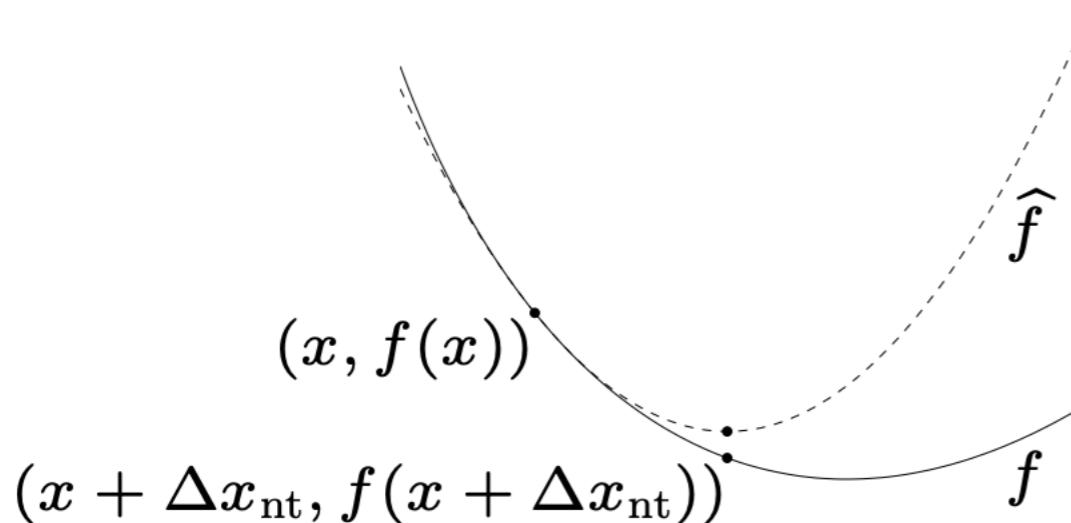
interpretations

- $x + \Delta x_{\text{nt}}$ minimizes second order approximation

$$\hat{f}(x + v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

- $x + \Delta x_{\text{nt}}$ solves linearized optimality condition

$$\nabla f(x + v) \approx \nabla \hat{f}(x + v) = \nabla f(x) + \nabla^2 f(x)v = 0$$



Its also a cool dance move
that Newton did

Newton step

$$\Delta x_{\text{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

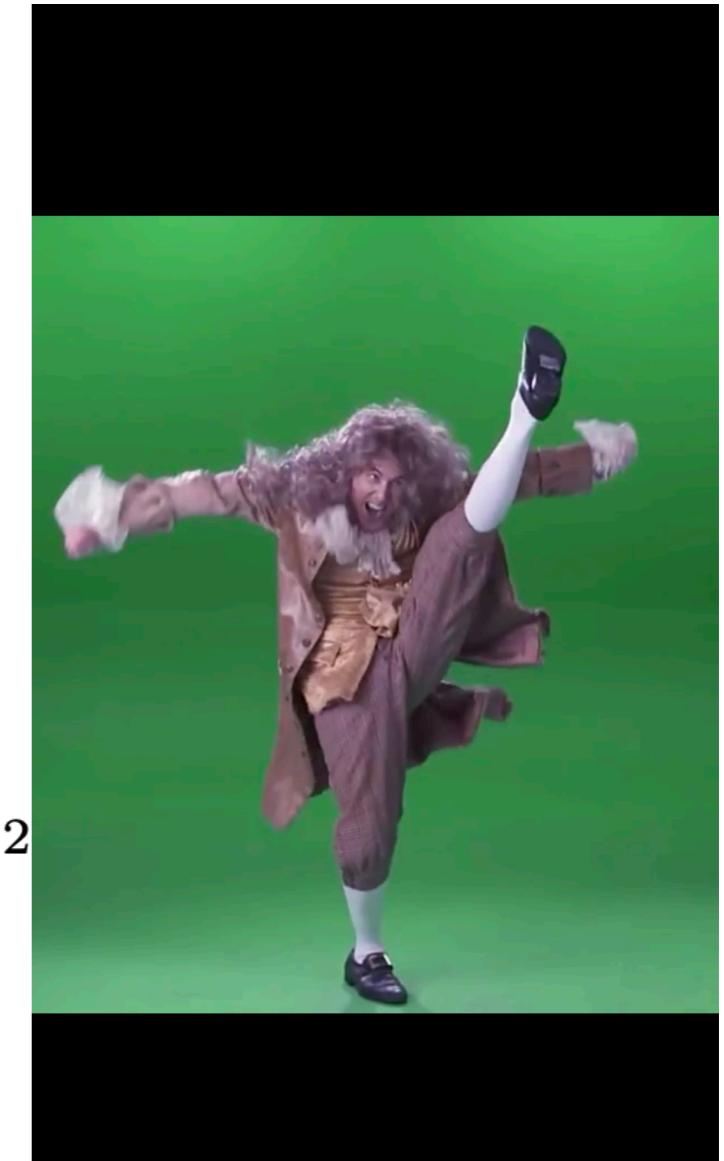
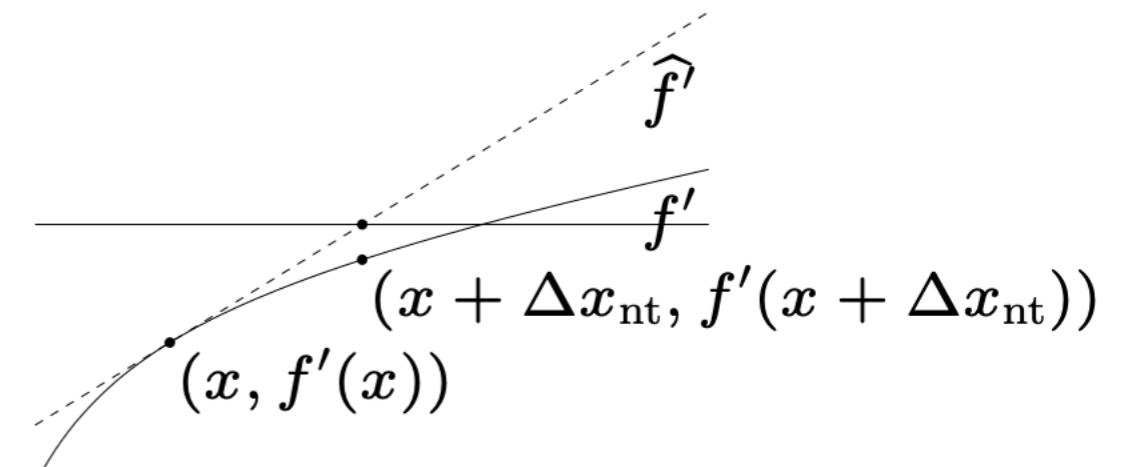
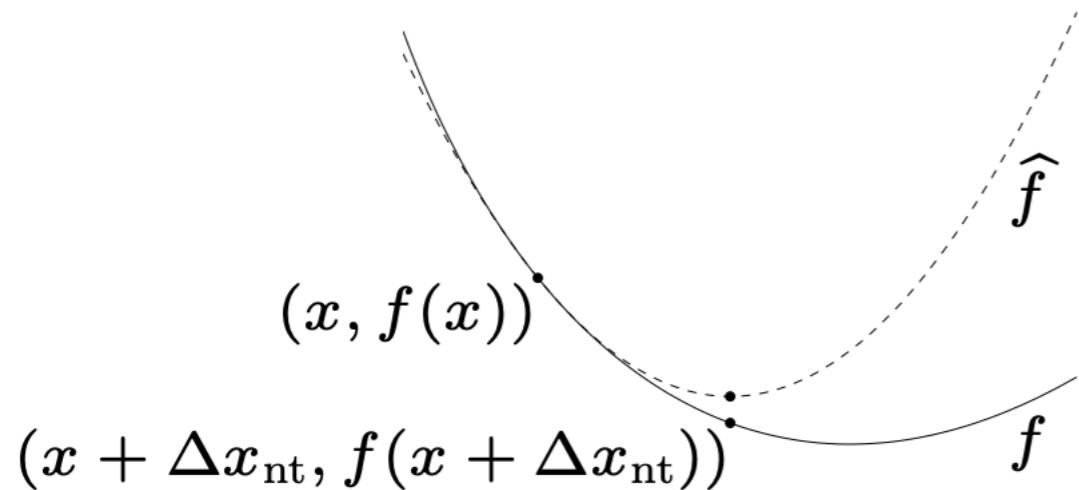
interpretations

- $x + \Delta x_{\text{nt}}$ minimizes second order approximation

$$\hat{f}(x + v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

- $x + \Delta x_{\text{nt}}$ solves linearized optimality condition

$$\nabla f(x + v) \approx \nabla \hat{f}(x + v) = \nabla f(x) + \nabla^2 f(x)v = 0$$



Uncertainty

What are we fitting?

- Lets say a fitted parameter has a distribution
 - We would like to quote the uncertainty on that distribution
 - Traditionally we do this by computing the standard deviation
 - Standard Deviation $\sigma = \sqrt{Var[\theta]}$
 - Standard deviation is just a number
 - It does not reflect what the actual distribution was about

How do we deal with unc?

- Our uncertainties have the same standard deviation
 - You may know there is an implicit assumption its Gaussian

