# Psych 253
# Advanced Statistical Modeling

## Review of Differential Calculus

Daniel Yamins
Wu Tsai Neurosciences Institute
Departments of Psychology and Computer Science
Stanford Artificial Intelligence Laboratory
Stanford University

Russ Poldrack
Department of Psychology
Stanford University

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***.

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^2$

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^2$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x}$$ (Def'n of derivative)

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^2$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x} \qquad \text{(Def'n of derivative)}$$

$$= \lim_{\Delta x \to 0} \frac{x^2 + 2\Delta x \cdot x + (\Delta x)^2 - x^2}{\Delta x} \qquad \text{(FOIL it out)}$$

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^2$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^2 - x^2}{\Delta x}$$ (Def'n of derivative)

$$= \lim_{\Delta x \to 0} \frac{x^2 + 2\Delta x \cdot x + (\Delta x)^2 - x^2}{\Delta x}$$ (FOIL it out)

$$= \lim_{\Delta x \to 0} \frac{2\Delta x \cdot x + (\Delta x)^2}{\Delta x}$$

(cancel leading term)

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^2$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x}$$ (Def'n of derivative)

$$= \lim_{\Delta x \to 0} \frac{x^2 + 2\Delta x \cdot x + (\Delta x)^2 - x^2}{\Delta x}$$ (FOIL it out)

$$= \lim_{\Delta x \to 0} \frac{2\Delta x \cdot x + (\Delta x)^2}{\Delta x} = \lim_{\Delta x \to 0} 2x + \Delta x = \boxed{2x}$$

(cancel leading term) (divide out **Δx**) (take limit)

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

Ex: $F(x) = x^n$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x + \Delta x)^n - x^n}{\Delta x} \quad \text{(Def'n of derivative)}$$

$$= \lim_{\Delta x \to 0} \frac{\sum_{i=0}^{n} \binom{n}{i} x^{n-i} (\Delta x)^i - x^n}{\Delta x} \quad \text{(binomial thm.)}$$

$$= \lim_{\Delta x \to 0} \frac{x^n + n x^{n-1} \cdot (\Delta x) + \sum_{i=2}^{n} \binom{n}{i} x^{n-i} (\Delta x)^i - x^n}{\Delta x}$$

$$= n x^{n-1} + \lim_{\Delta x \to 0} \sum_{i=2}^{n} \binom{n}{i} x^{n-i} (\Delta x)^{i-1} = \boxed{n x^{n-1}}$$

# Definition of Derivative

Let ***F(x)*** be a real-valued function of one real variable ***x***. The derivative is defined as:

$$F'(x) = \frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}$$

NB: these three steps, in this specific computation, represent Newton's most important mathematical contributions, shifting the course of history after ever

Ex: $F(x) = x^n$

$$F'(x) = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^n - x^n}{\Delta x} \quad \text{(Def'n of derivative)}$$

$$= \lim_{\Delta x \to 0} \frac{\sum_{i=0}^{n} \binom{n}{i} x^{n-i} (\Delta x)^i - x^n}{\Delta x} \quad \text{(binomial thm.)}$$

$$= \lim_{\Delta x \to 0} \frac{x^n + n x^{n-1} \cdot (\Delta x) + \sum_{i=2}^{n} \binom{n}{i} x^{n-i} (\Delta x)^i - x^n}{\Delta x}$$

$$= n x^{n-1} + \lim_{\Delta x \to 0} \sum_{i=2}^{n} \binom{n}{i} x^{n-i} (\Delta x)^{i-1} = \boxed{n x^{n-1}}$$

# Interpretation as Local Slope

***F'(x)*** is the slope of the line that is tangent to the graph of ***F(x)*** at ***x***:

Suppose that line is:

$$y = mx + b$$

Then:

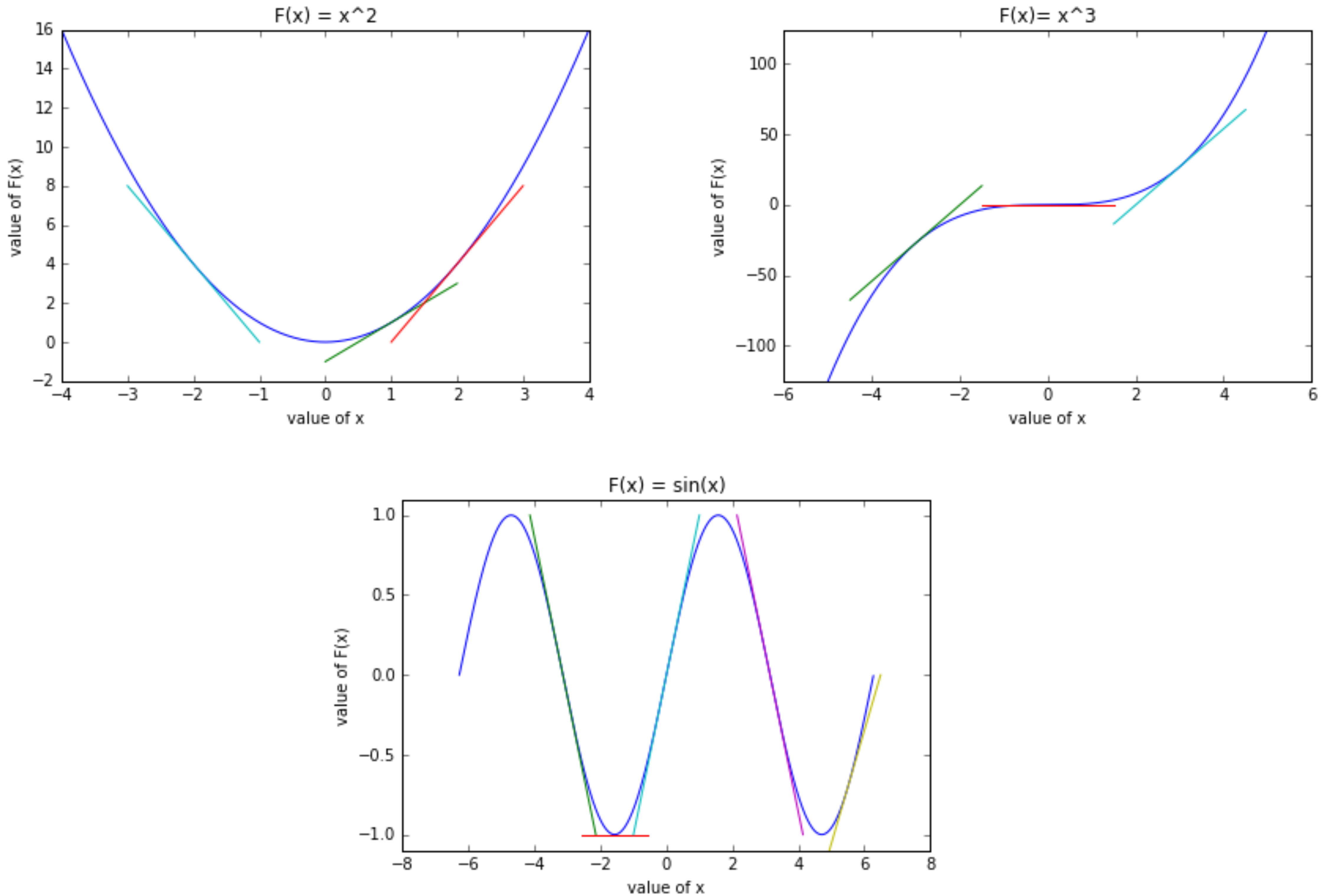
$$F(x) = F'(x)x + b$$

So:

$$b = F(x) - F'(x)x$$

# Interpretation as Local Slope

***F'(x)*** is the slope of the line that is tangent to the graph of ***F(x)*** at ***x***:



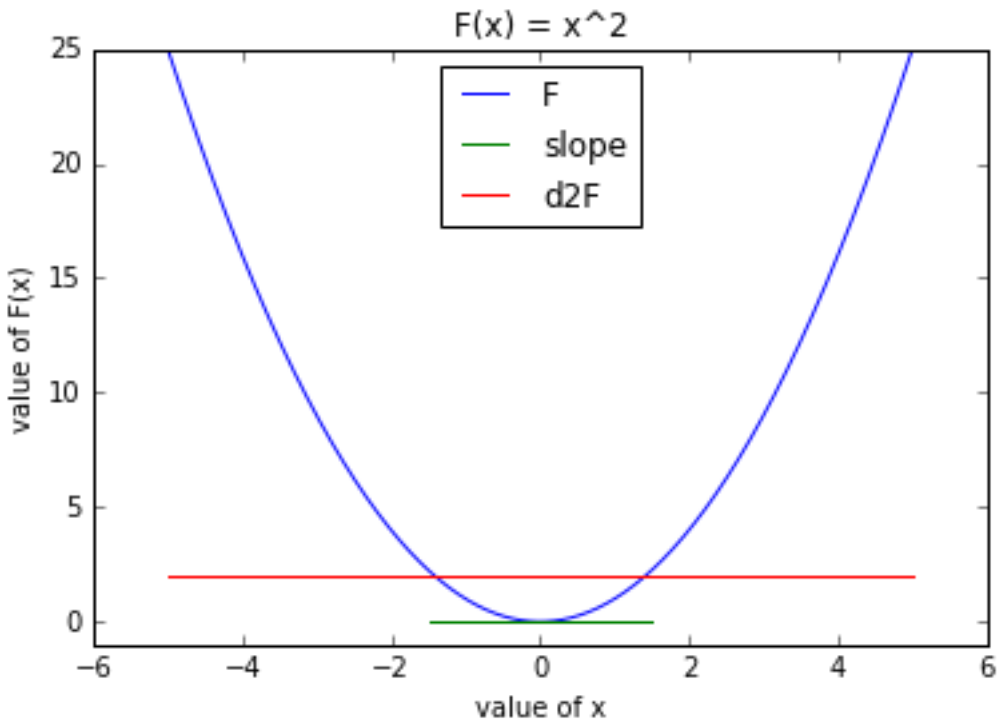
# Higher Derivatives

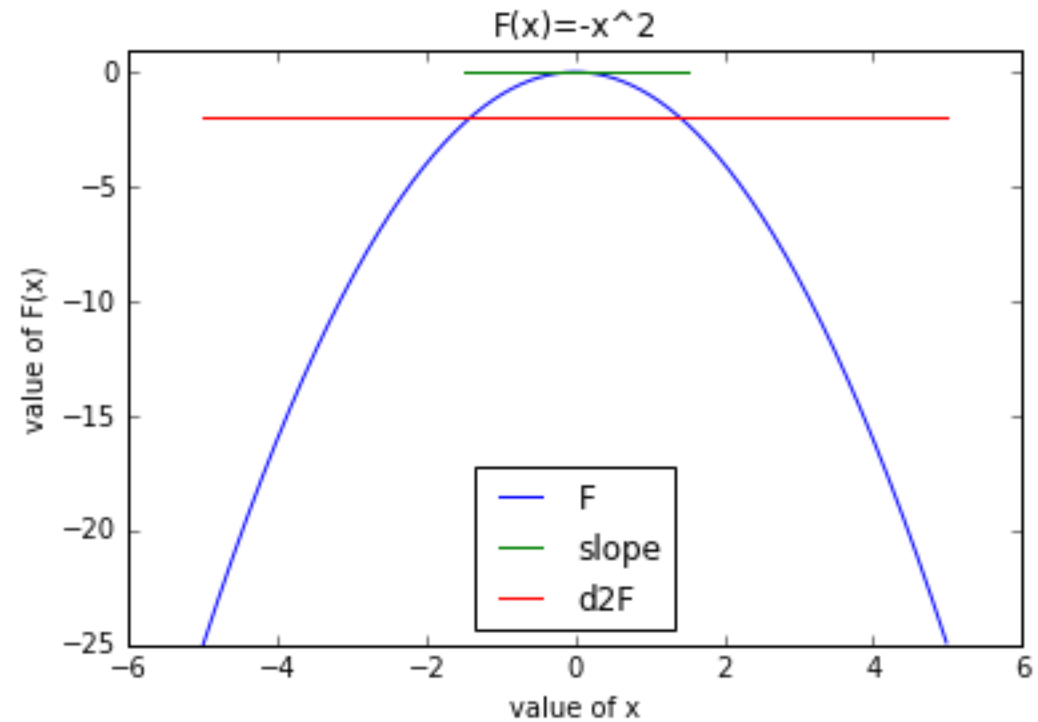
The nth derivative is the derivative of the (n-1)st derivative:

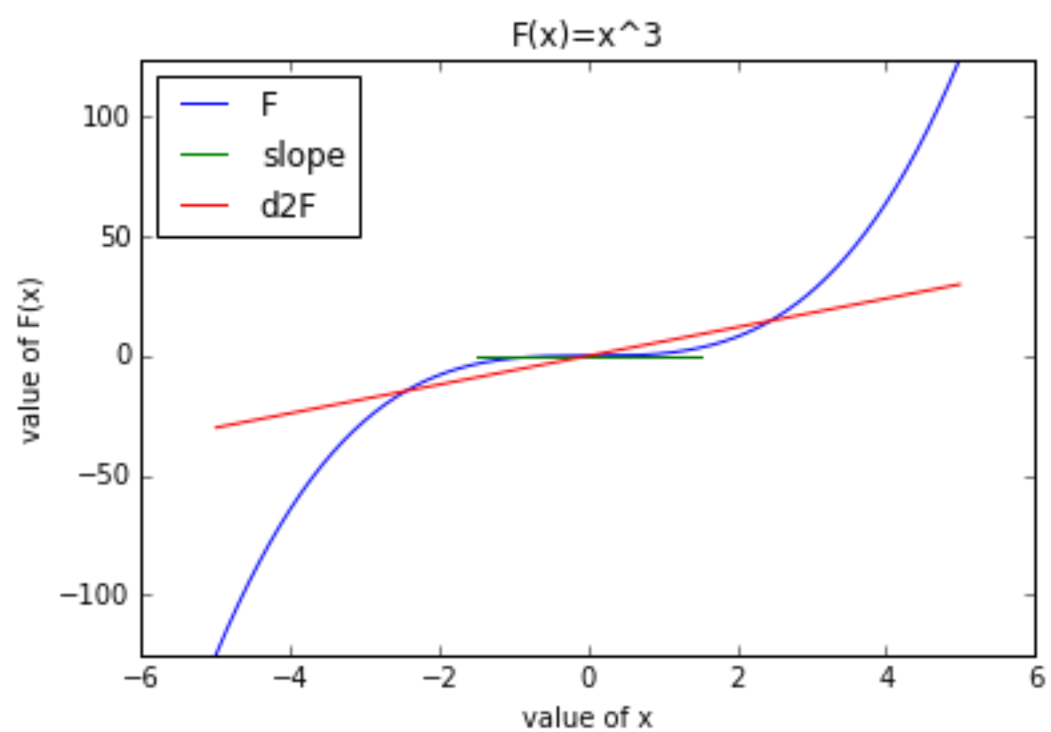
$$F^{(n)}(x) = \frac{d^n F}{dx^n} = \frac{d}{dx}(F^{(n-1)}(x))$$

# Higher Derivatives

E.g., second derivatives:



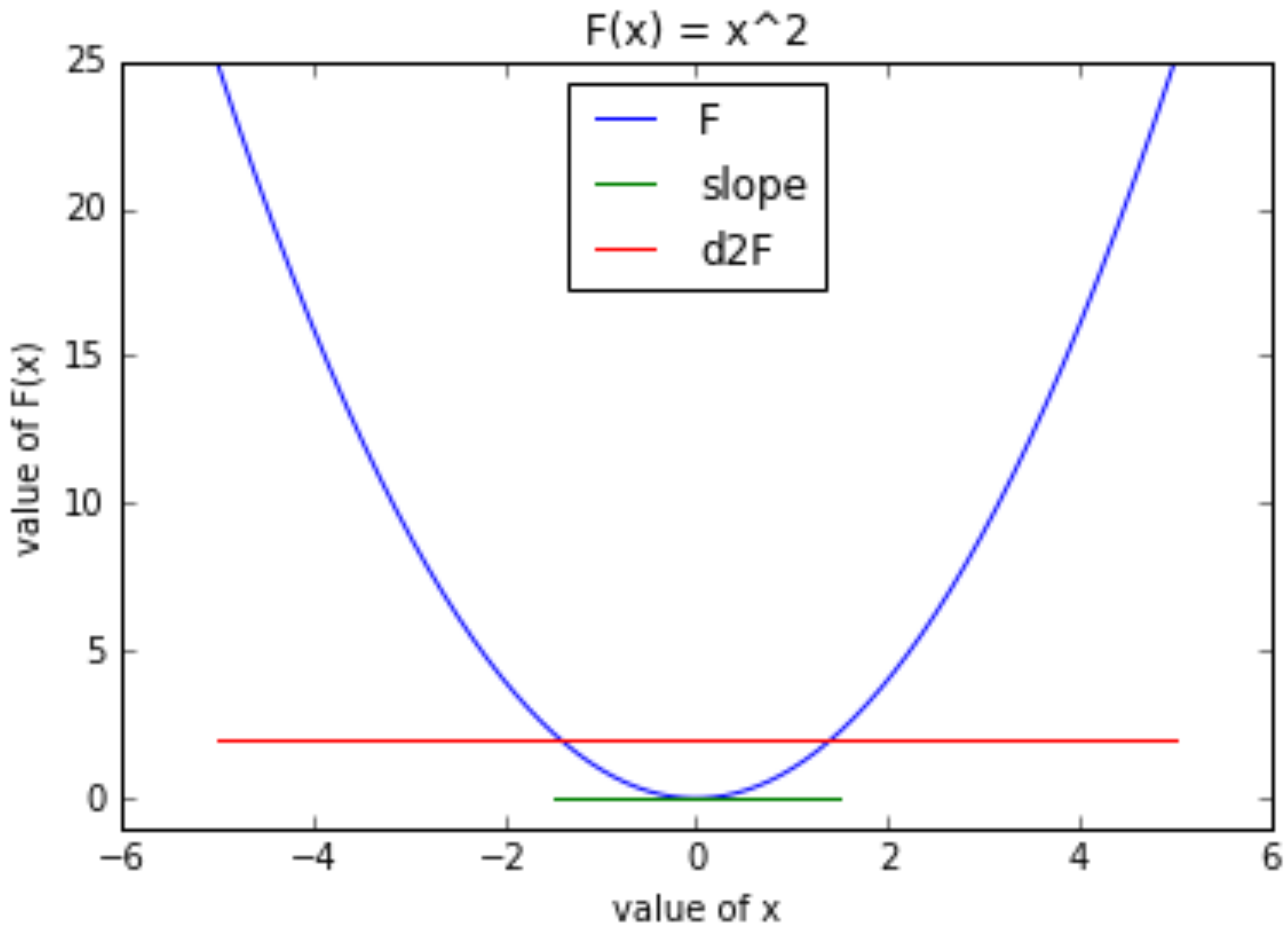



# Critical Points and Local Optima

Critical points are those ***x**** such that:

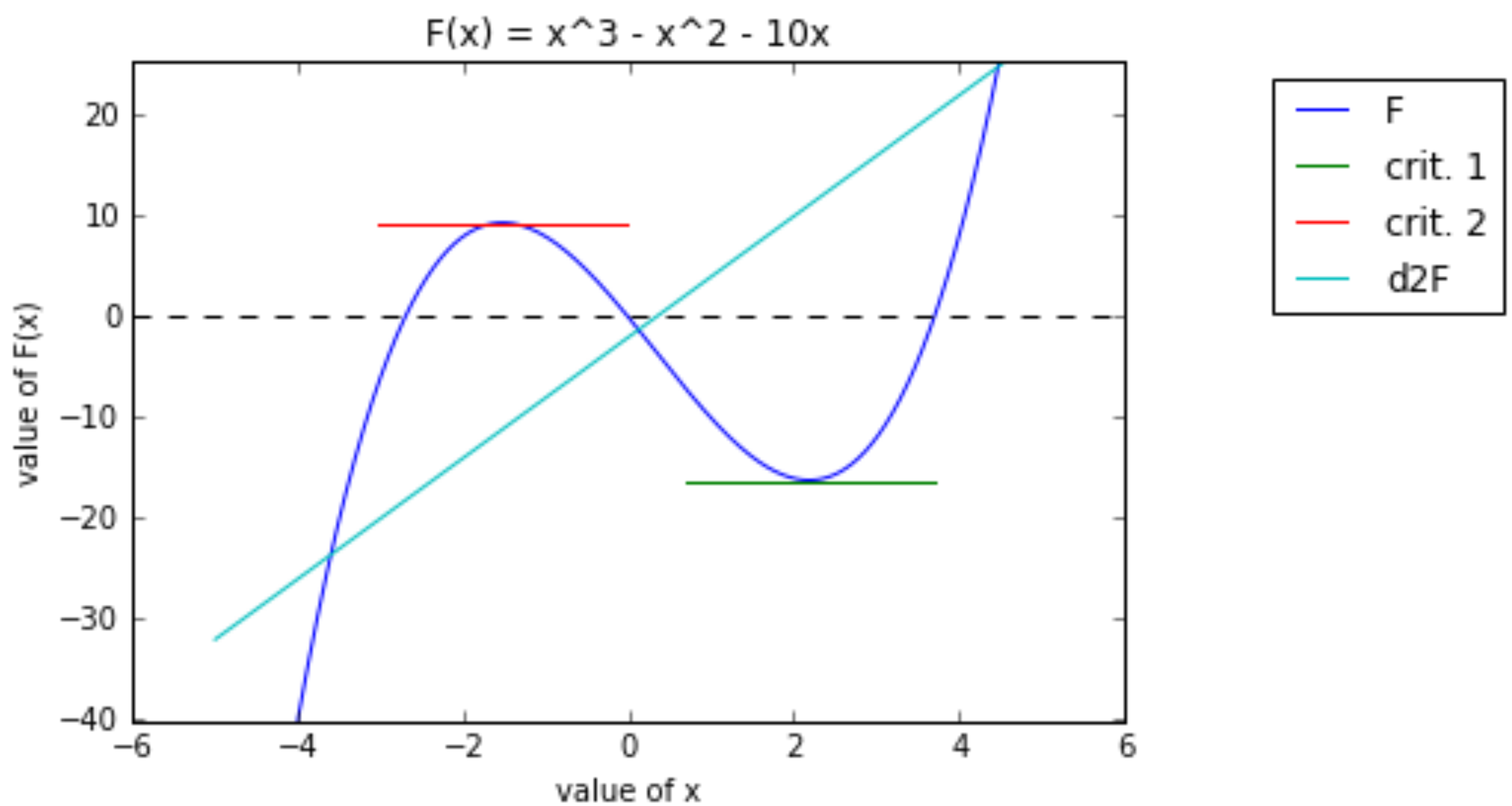
$$\left.\frac{dF}{dx}\right|_{x=x^*} = 0$$



# Critical Points and Local Optima

Critical points are those ***x**** such that:

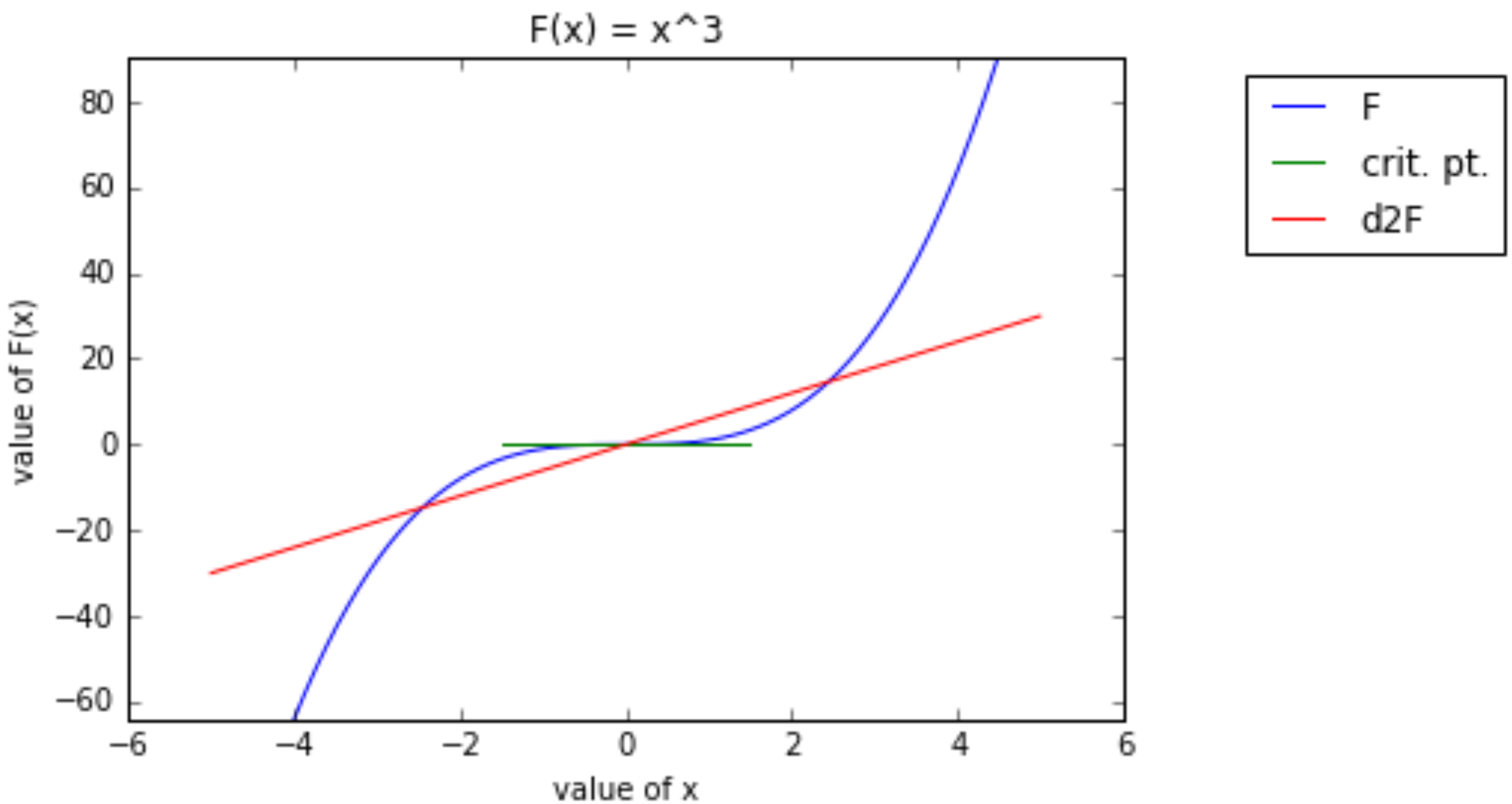
$$\left.\frac{dF}{dx}\right|_{x=x^*} = 0$$



# Critical Points and Local Optima

Critical points are those ***x**** such that:

$$\left.\frac{dF}{dx}\right|_{x=x^*} = 0$$



# Critical Points and Local Optima

Critical points are those ***x**** such that:

$$\left.\frac{dF}{dx}\right|_{x=x^*} = 0$$

Second derivative test:

$$\text{critical point } x^* = \left\{ \begin{array}{ll} \text{local maximum,} & \text{if } d^2F/dx^2 < 0 \\ \text{local minimum,} & \text{if } d^2F/dx^2 > 0 \\ \text{saddle point (maybe),} & \text{if } d^2F/dx^2 = 0 \end{array} \right\}$$

# Taylor Series

Let's say we wanted to approximate a function ***F*** around a point ***x0*** as a series in **Δ*x***:

$$F(x_0 + \Delta x) = a_0 + a_1 \Delta x + a_2 [\Delta x]^2 + a_3 [\Delta x]^3 + \ldots$$

# Taylor Series

Let's say we wanted to approximate a function ***F*** around a point ***x0*** as a series in ***Δx***:

$$F(x_0 + \Delta x) = a_0 + a_1 \Delta x + a_2 [\Delta x]^2 + a_3 [\Delta x]^3 + \ldots$$

$$= \sum_{i=0}^{\infty} a_i [\Delta x]^i$$

# Taylor Series

Let's say we wanted to approximate a function ***F*** around a point ***x0*** as a series in **Δ*x***:

constant term

$$F(x_0 + \Delta x) = \boxed{a_0} + a_1 \Delta x + a_2 [\Delta x]^2 + a_3 [\Delta x]^3 + \ldots$$

$$= \sum_{i=0}^{\infty} a_i [\Delta x]^i$$

# Taylor Series

Let's say we wanted to approximate a function ***F*** around a point ***x0*** as a series in **Δx**:

$$F(x_0 + \Delta x) = \overbrace{\boxed{a_0}}^{\text{constant term}} + \underbrace{\boxed{a_1 \Delta x}}_{\text{linear team}} + \overbrace{\boxed{a_2 [\Delta x]^2}}^{\text{quadratic term}} + \underbrace{\boxed{a_3 [\Delta x]^3}}_{\text{cubic term}} + \dots$$

$$= \sum_{i=0}^{\infty} a_i [\Delta x]^i$$

First of all: $\Delta x = 0 \rightarrow a_0 = F(x_0)$

Then: $\dfrac{dF}{d\Delta x}(x_0 + \Delta x) = a_1 + 2a_2 \Delta x + \dots = \sum_{i=1}^{\infty} i \cdot a_i [\Delta x]^{i-1}$

So: $\Delta x = 0 \rightarrow a_1 = \left.\dfrac{dF}{dx}\right|_{x=x_0}$

# Taylor Series

But if $$\frac{dF}{d\Delta x}(x_0 + \Delta x) = a_1 + 2a_2\Delta x + \ldots = \sum_{i=1}^{\infty} i \cdot a_i [\Delta x]^{i-1}$$

then $$\frac{d^2F}{d\Delta x^2}(x_0 + \Delta x) = 2a_2 + 6a_3\Delta x + \ldots = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i [\Delta x]^{i-2}$$

# Taylor Series

But if $$\frac{dF}{d\Delta x}(x_0 + \Delta x) = a_1 + 2a_2\Delta x + \ldots = \sum_{i=1}^{\infty} i \cdot a_i[\Delta x]^{i-1}$$

then $$\frac{d^2F}{d\Delta x^2}(x_0 + \Delta x) = 2a_2 + 6a_3\Delta x + \ldots = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i[\Delta x]^{i-2}$$

thus again, $$\Delta x = 0 \rightarrow a_2 = \frac{1}{2}\frac{d^2F}{dx^2}\bigg|_{x=x_0}$$

# Taylor Series

But if $$\frac{dF}{d\Delta x}(x_0 + \Delta x) = a_1 + 2a_2\Delta x + \ldots = \sum_{i=1}^{\infty} i \cdot a_i[\Delta x]^{i-1}$$

then $$\frac{d^2F}{d\Delta x^2}(x_0 + \Delta x) = 2a_2 + 6a_3\Delta x + \ldots = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i[\Delta x]^{i-2}$$

thus again, $$\Delta x = 0 \rightarrow a_2 = \frac{1}{2}\frac{d^2F}{dx^2}\bigg|_{x=x_0}$$

In general:

$$k \cdot (k-1) \ldots 2 \cdot 1 \cdot a_k = \frac{d^kF}{dx^k}\bigg|_{x=x_0}$$

# Taylor Series

But if $$\frac{dF}{d\Delta x}(x_0 + \Delta x) = a_1 + 2a_2\Delta x + \ldots = \sum_{i=1}^{\infty} i \cdot a_i [\Delta x]^{i-1}$$

then $$\frac{d^2F}{d\Delta x^2}(x_0 + \Delta x) = 2a_2 + 6a_3\Delta x + \ldots = \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i [\Delta x]^{i-2}$$

thus again, $$\Delta x = 0 \rightarrow a_2 = \frac{1}{2}\frac{d^2F}{dx^2}\bigg|_{x=x_0}$$

In general:

$$k \cdot (k-1) \ldots 2 \cdot 1 \cdot a_k = \frac{d^kF}{dx^k}\bigg|_{x=x_0}$$

Or:

$$a_k = \frac{1}{k!}\frac{d^kF}{dx^k}\bigg|_{x=x_0}$$

# Taylor Series

Let's say we wanted to approximate a function ***F*** around a point ***x0*** as a series in **Δ*x***:

$$F(x_0 + \Delta x) = a_0 + a_1 \Delta x + a_2 [\Delta x]^2 + a_3 [\Delta x]^3 + \ldots$$

$$= \sum_{k=0}^{\infty} a_k [\Delta x]^k$$

$$= \boxed{\sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{d^k F}{dx^k} \right|_{x=x_0} \cdot [\Delta x]^k}$$

Taylor Series Approximation

# Taylor Series

Let's calculate the derivative of the sine function.

$$F(x) = \sin(x)$$

# Taylor Series

Let's calculate the derivative of the sine function.

$$F(x) = \sin(x)$$

$$F'(x) = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin(x)}{\Delta x}$$

Let's calculate the derivative of the sine function.

$$F(x) = \sin(x)$$

$$F'(x) = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\sin(x)\cos(\Delta x) + \cos(x)\sin(\Delta x) - \sin(x)}{\Delta x}$$

# Taylor Series

Let's calculate the derivative of the sine function.

$$F(x) = \sin(x)$$

$$F'(x) = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\sin(x)\cos(\Delta x) + \cos(x)\sin(\Delta x) - \sin(x)}{\Delta x}$$

$$= \sin(x) \lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} + \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

Let's calculate the derivative of the sine function.

$$F(x) = \sin(x)$$

$$F'(x) = \lim_{\Delta x \to 0} \frac{\sin(x + \Delta x) - \sin(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\sin(x)\cos(\Delta x) + \cos(x)\sin(\Delta x) - \sin(x)}{\Delta x}$$

$$= \sin(x) \boxed{\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x}} + \cos(x) \boxed{\lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}}$$

So now we need to calculate these two limits

# Taylor Series

We'll use the Taylor Series concept to calculate the first limit.

# Taylor Series

Cosine's Taylor series around 0 is:

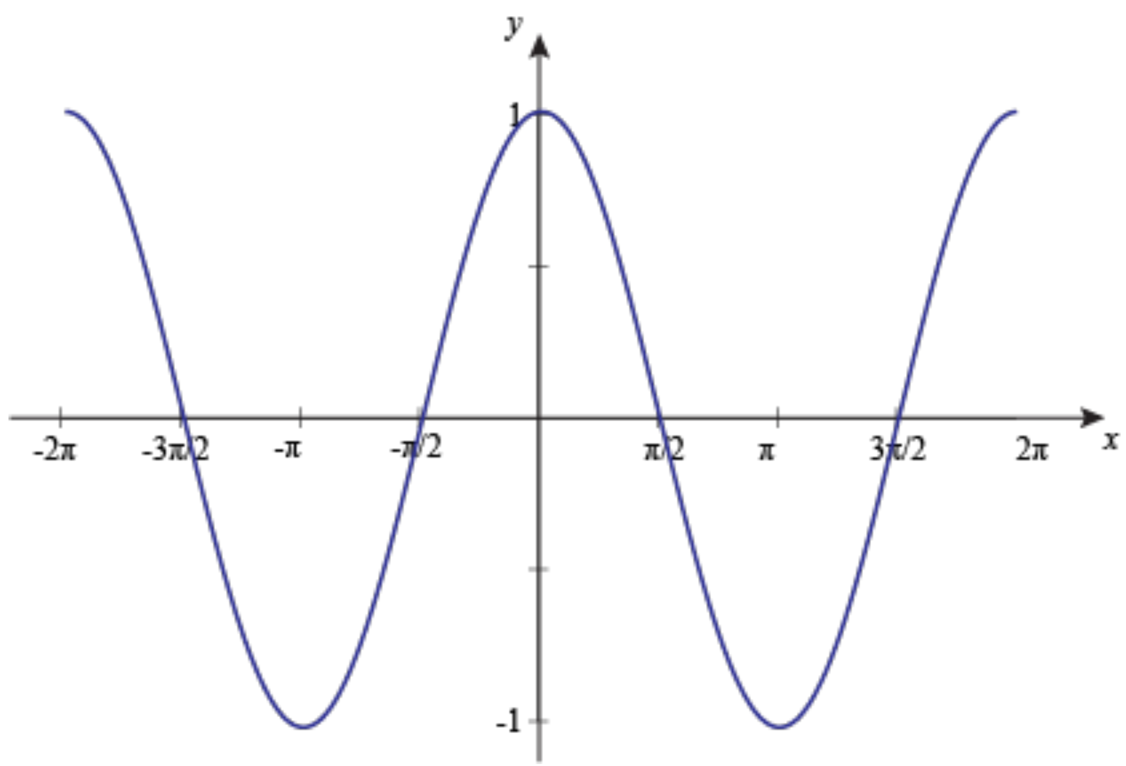
$$\cos(\Delta x) = \cos(0) + a_1 \cos'(0)\Delta x + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

# Taylor Series

Cosine's Taylor series around 0 is:

$$\cos(\Delta x) = \cos(0) + a_1 \cos'(0)\Delta x + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

Now, ***cos(x)*** is an even function around 0:



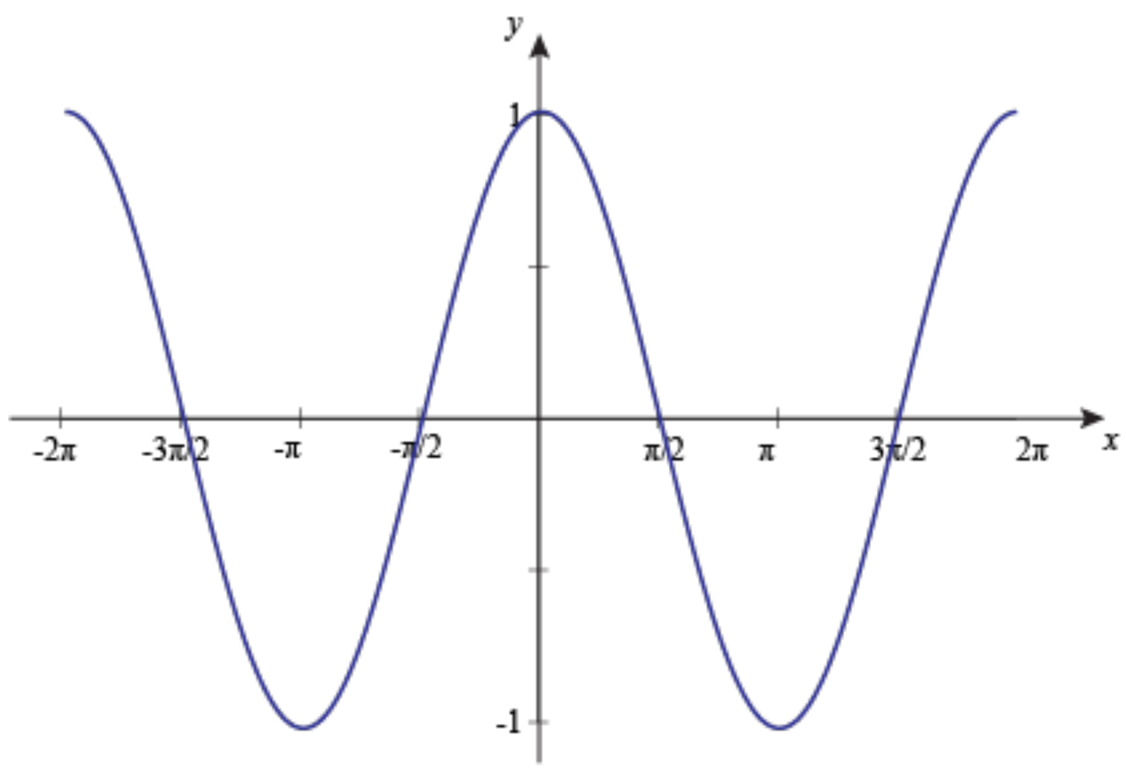

$$\cos(x) = \cos(-x)$$

# Taylor Series

Cosine's Taylor series around 0 is:

$$\cos(\Delta x) = \cos(0) + a_1 \cos'(0)\Delta x + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

Now, ***cos(x)*** is an even function around 0:




$$\cos(x) = \cos(-x)$$

Thus:

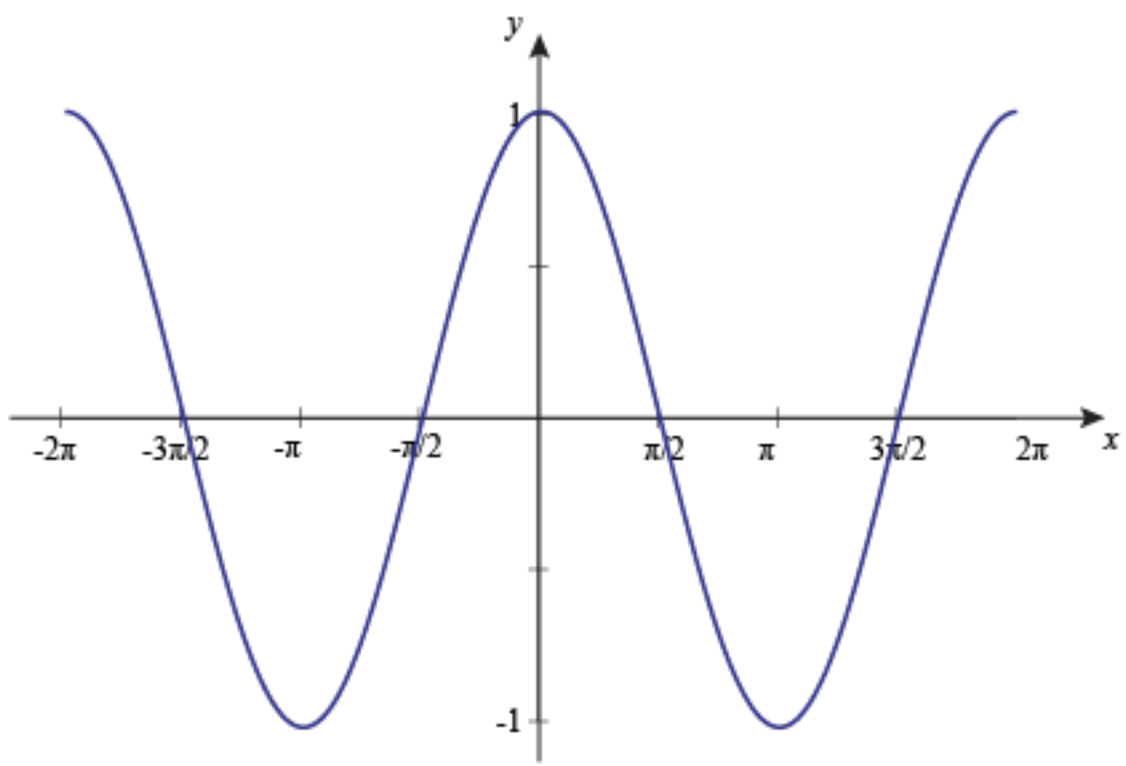
$$a_1 = 0$$

# Taylor Series

Cosine's Taylor series around 0 is:

$$\cos(\Delta x) = \cos(0) + a_1 \cos'(0)\Delta x + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

Now, ***cos(x)*** is an even function around 0:




$$\cos(x) = \cos(-x)$$

Thus:

$$a_1 = 0$$

Actually:

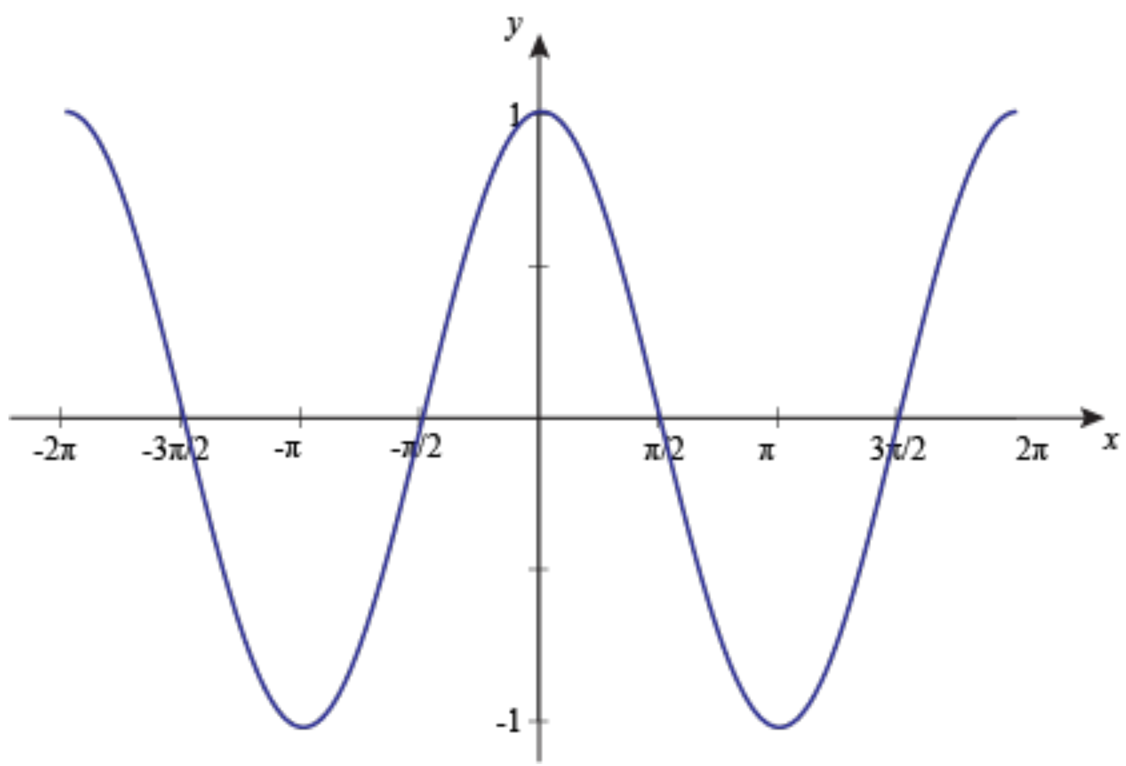
$$a_i = 0 \text{ all } i \text{ odd}$$

# Taylor Series

Cosine's Taylor series around 0 is:

$$\cos(\Delta x) = \cos(0) + a_1 \cos'(0)\Delta x + a_2 \cos''(0)(\Delta x)^2 + \dots$$

Now, ***cos(x)*** is an even function around 0:




$$\cos(x) = \cos(-x)$$

Thus:

$$a_1 = 0$$

Actually:

$$a_i = 0 \text{ all } i \text{ odd}$$

So cosine's Taylor series is:

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \dots$$

# Taylor Series

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a_2 \cos''(0)(\Delta x)^2}{\Delta x}$$

# Taylor Series

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a_2 \cos''(0)(\Delta x)^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} a_2 \cos''(0) \Delta x$$

# Taylor Series

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a_2 \cos''(0)(\Delta x)^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} a_2 \cos''(0)\Delta x$$

$$= 0$$

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a_2 \cos''(0)(\Delta x)^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} a_2 \cos''(0) \Delta x$$

$$= 0$$

So

$$\sin'(x) = \sin(x) \cancelto{= 0}{\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x}} + \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

# Taylor Series

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a_2 \cos''(0)(\Delta x)^2}{\Delta x}$$

$$= \lim_{\Delta x \to 0} a_2 \cos''(0) \Delta x$$

$$= 0$$

So

$$\sin'(x) = \sin(x) \cancelto{0}{\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x}} + \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

$$= \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \ldots$$

Observation: note that we didn't need to know cosine's full taylor series for this calculation — just that cosine's being an even fn meant that it's first order term disappeared … this is one reason ***why*** Taylor series are useful

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} \frac{a}{\phantom{x}}$$

$$= \lim_{\Delta x \to 0} a_2$$

$$= 0$$

So

$$\sin'(x) = \sin(x) \cancelto{0}{\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x}} + \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

$$= \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

# Taylor Series

Since

$$\cos(\Delta x) = 1 + a_2 \cos''(0)(\Delta x)^2 + \dots$$

we have:

$$\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x} = \lim_{\Delta x \to 0} a_2 \dots$$

$$= \lim_{\Delta x \to 0} a_2 \dots$$

$$= 0$$

Observation: note that we didn't need to know cosine's full taylor series for this calculation — just that cosine's being an even fn meant that it's first order term disappeared … this is one reason ***why*** Taylor series are useful

So

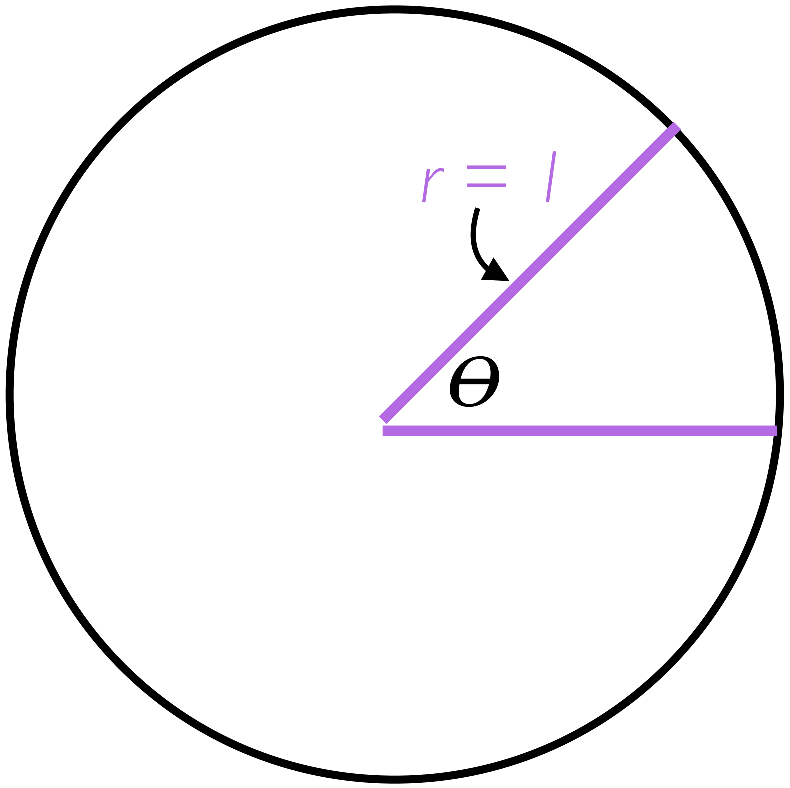
$$\sin'(x) = \sin(x) \underbrace{\cancel{\lim_{\Delta x \to 0} \frac{\cos(\Delta x) - 1}{\Delta x}}}_{= 0} + \cos(x) \lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}$$

$$= \cos(x) \underbrace{\boxed{\lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}}}_{= \sin'(0)}$$

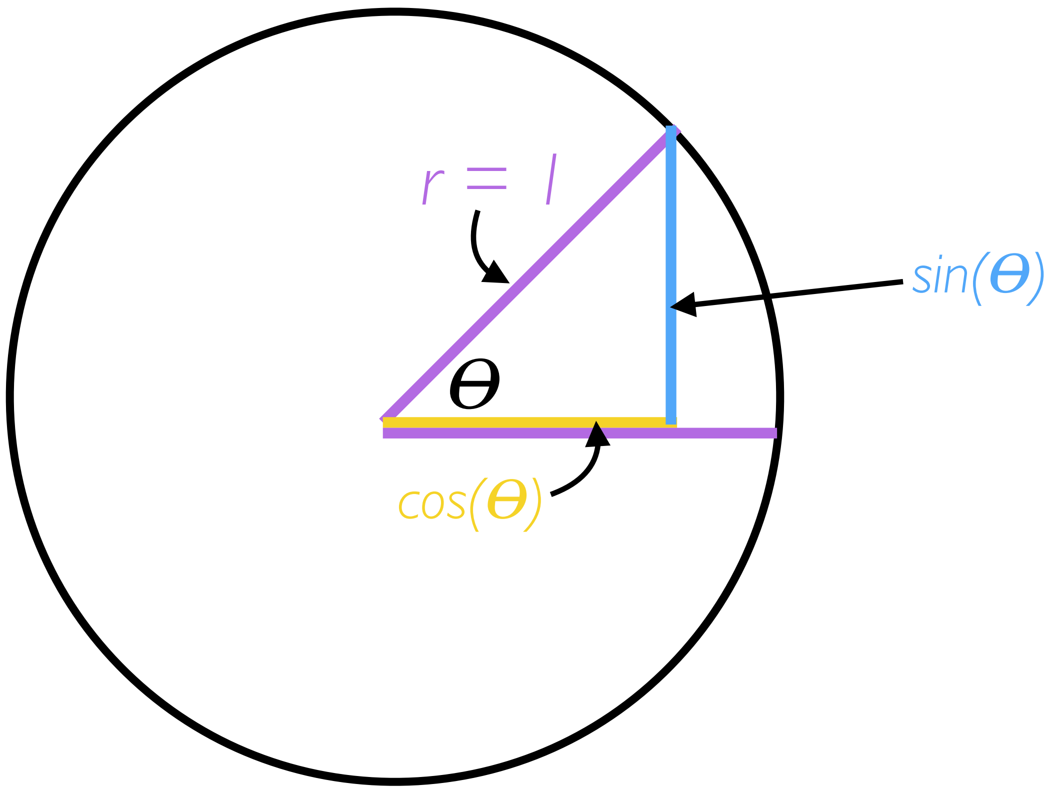
Ok, but what is ***this*** limit?

# Taylor Series

Assume circle is radius 1

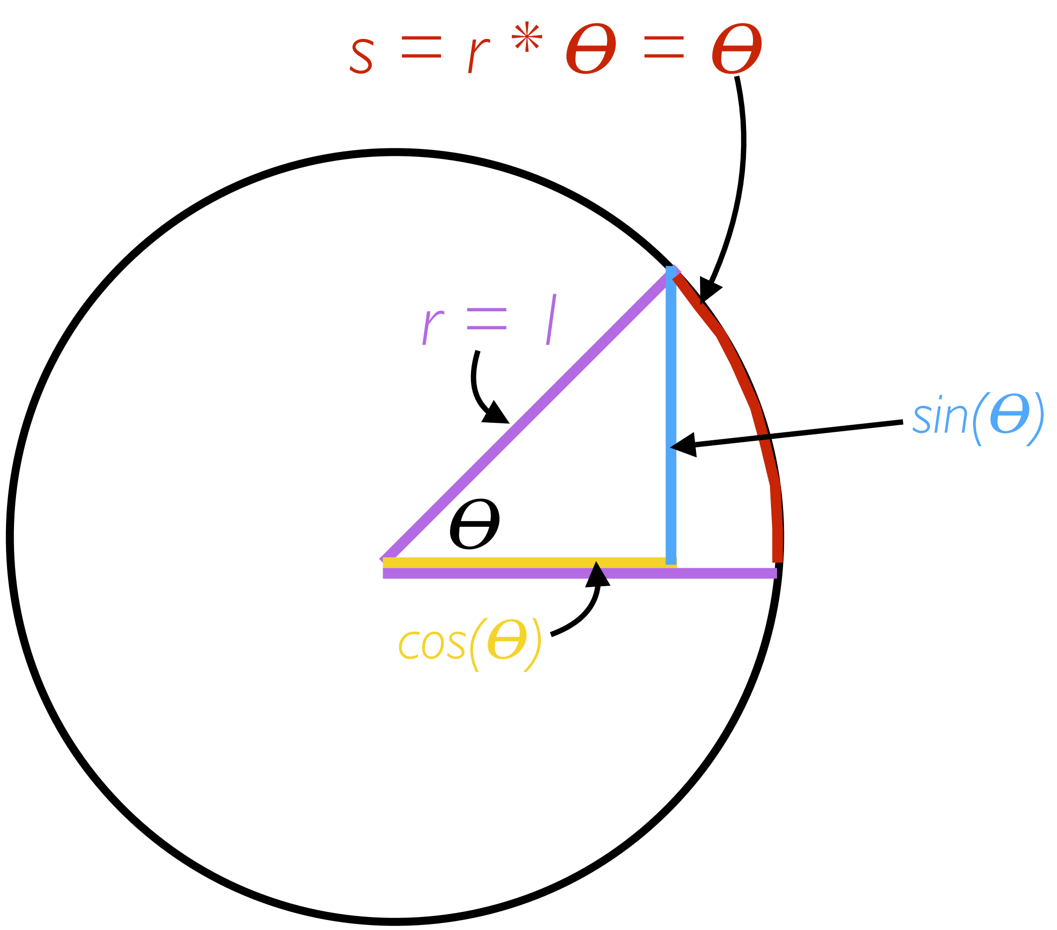


Assume circle is radius 1



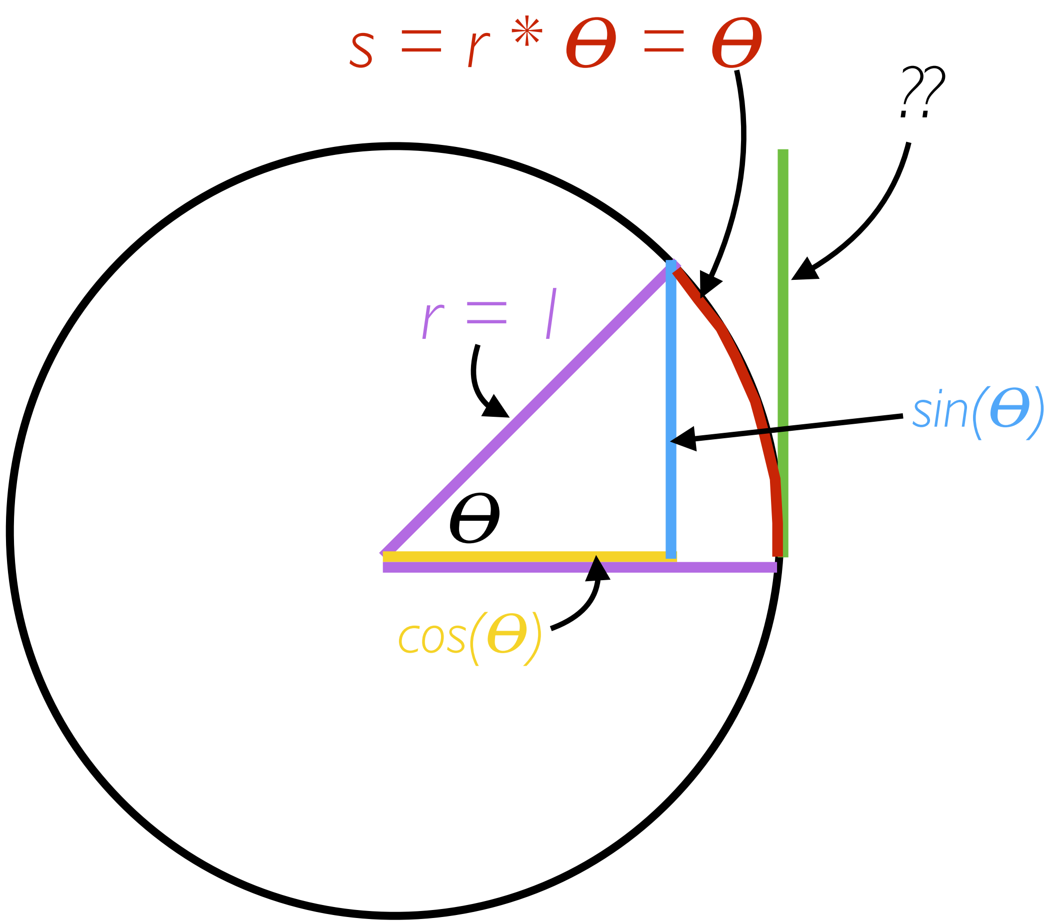
# Taylor Series

Assume circle is radius 1

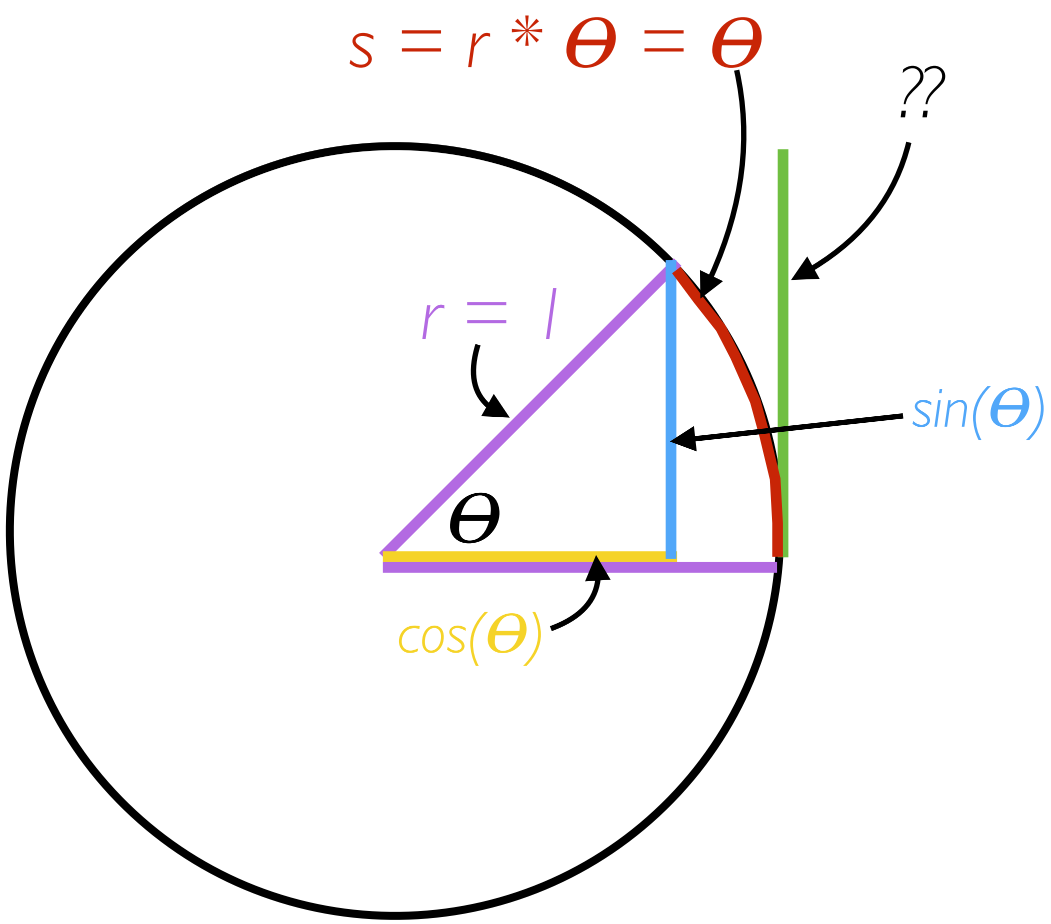


# Taylor Series

Assume circle is radius 1



Assume circle is radius 1



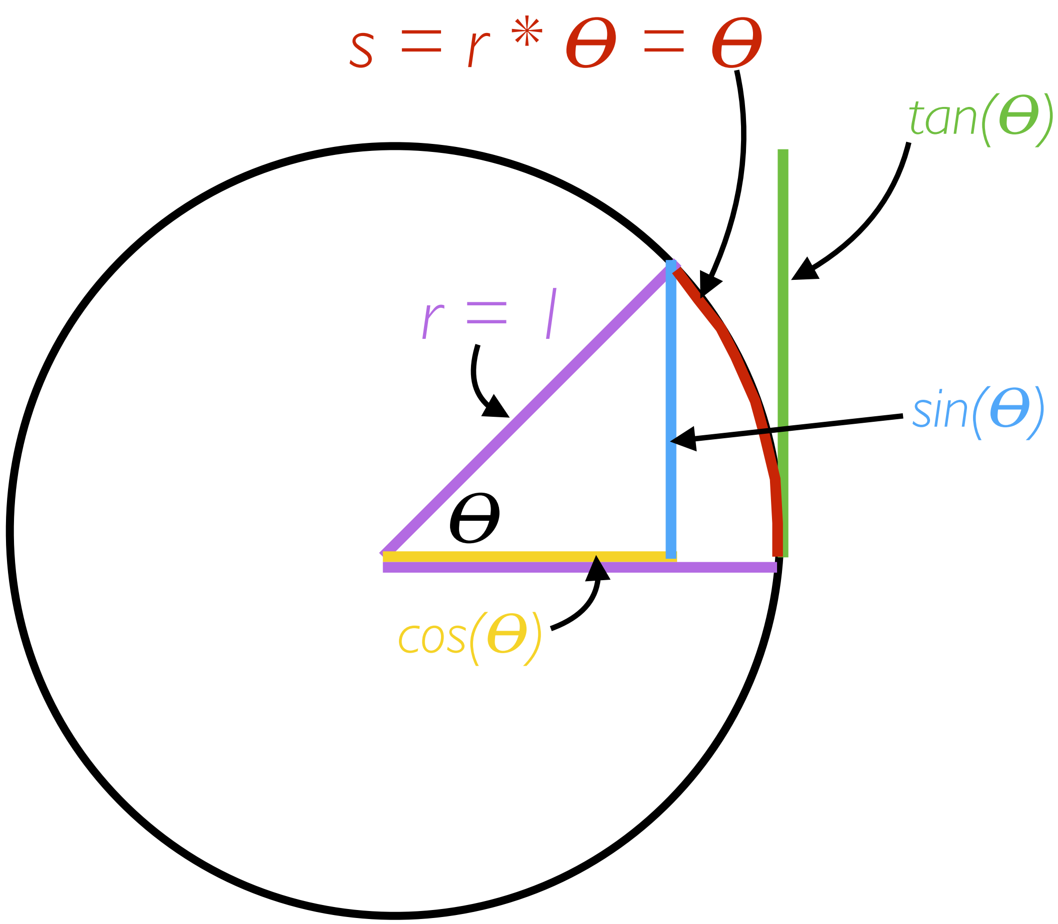

$$\frac{\sin(\theta)}{\cos(\theta)}$$ = ratio of blue to yellow = ratio of green to **r** = green

$$=$$

$$\tan(\theta)$$

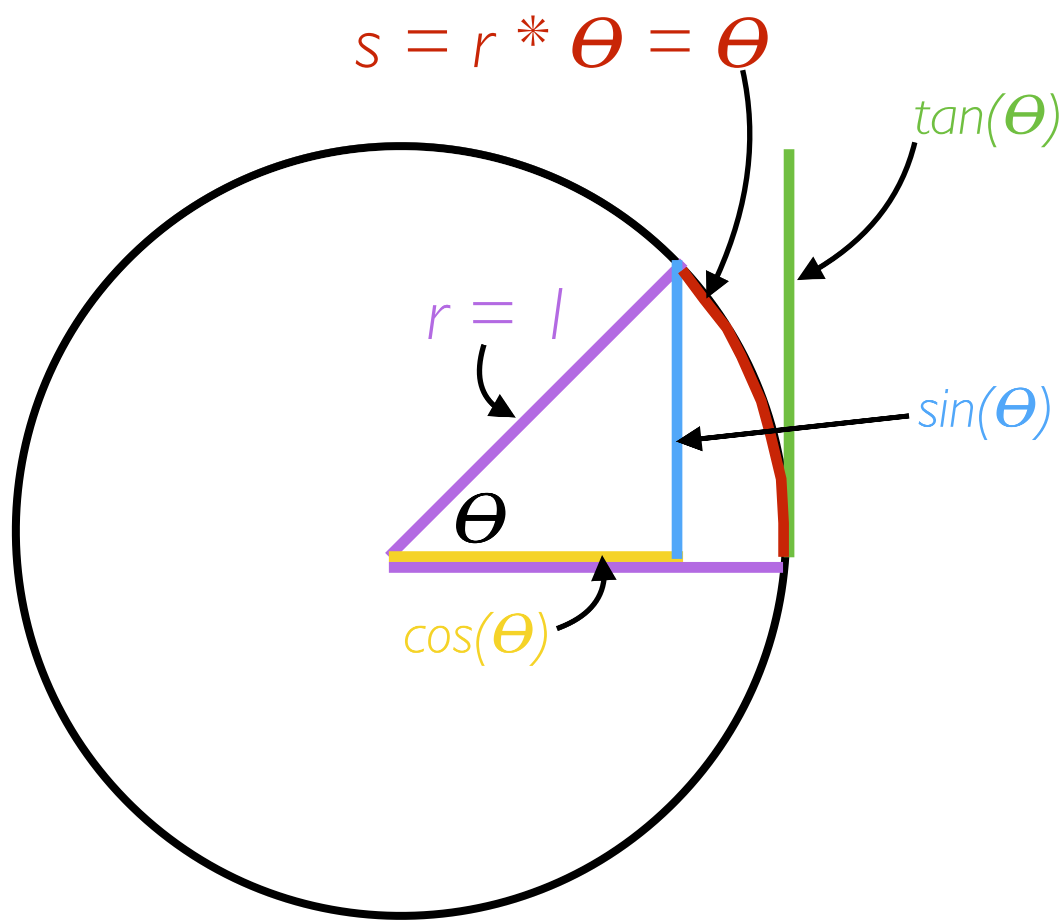
# Taylor Series

Assume circle is radius 1

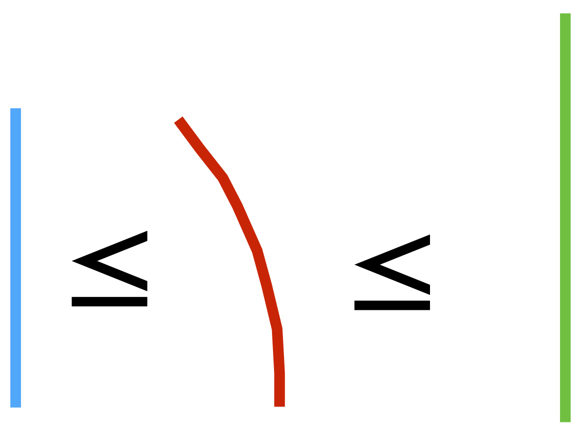


# Taylor Series

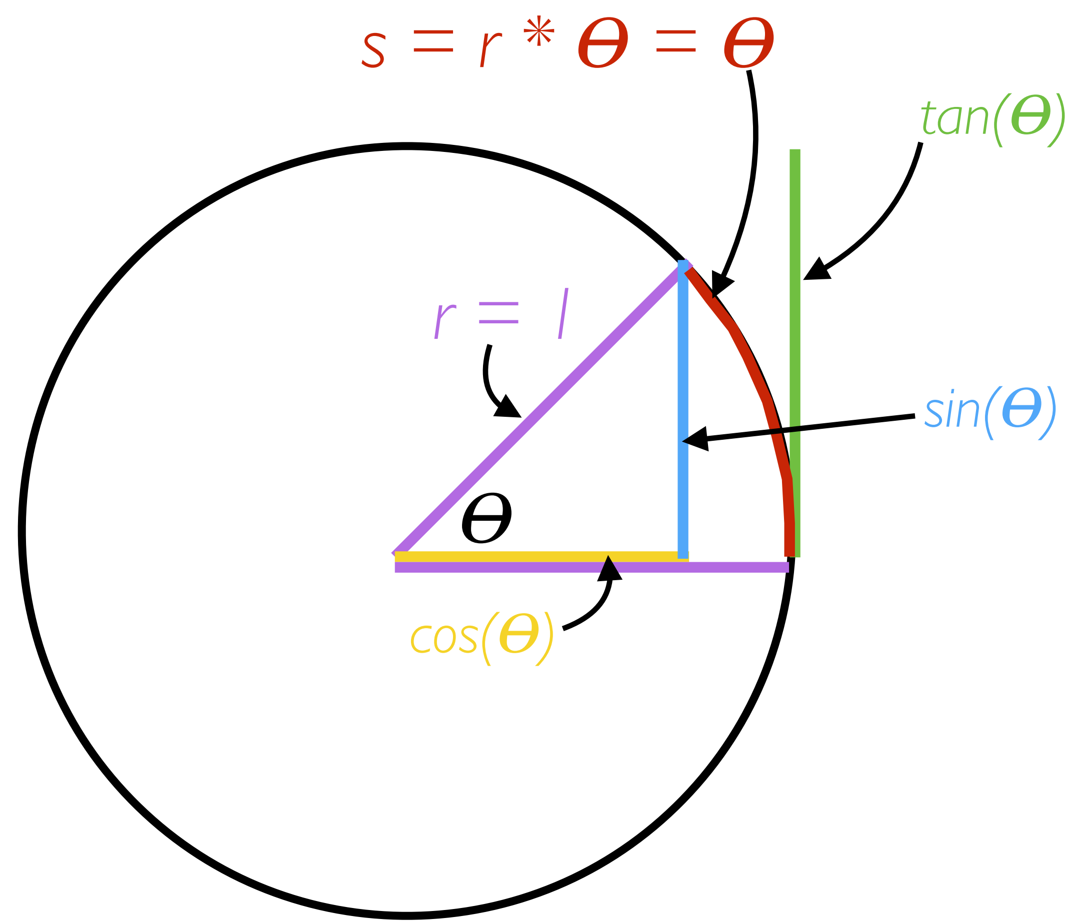
Assume circle is radius 1



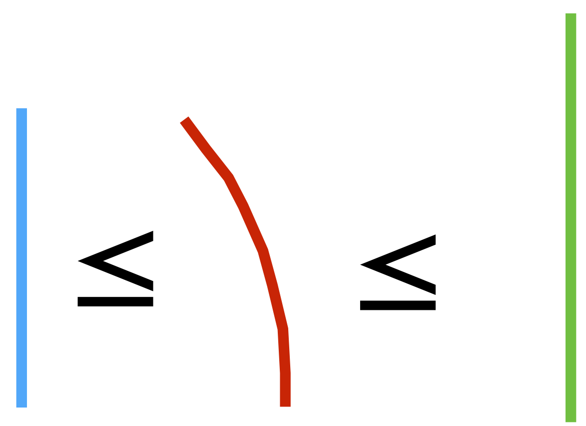

But obviously, the lengths:



Assume circle is radius 1




But obviously, the lengths:

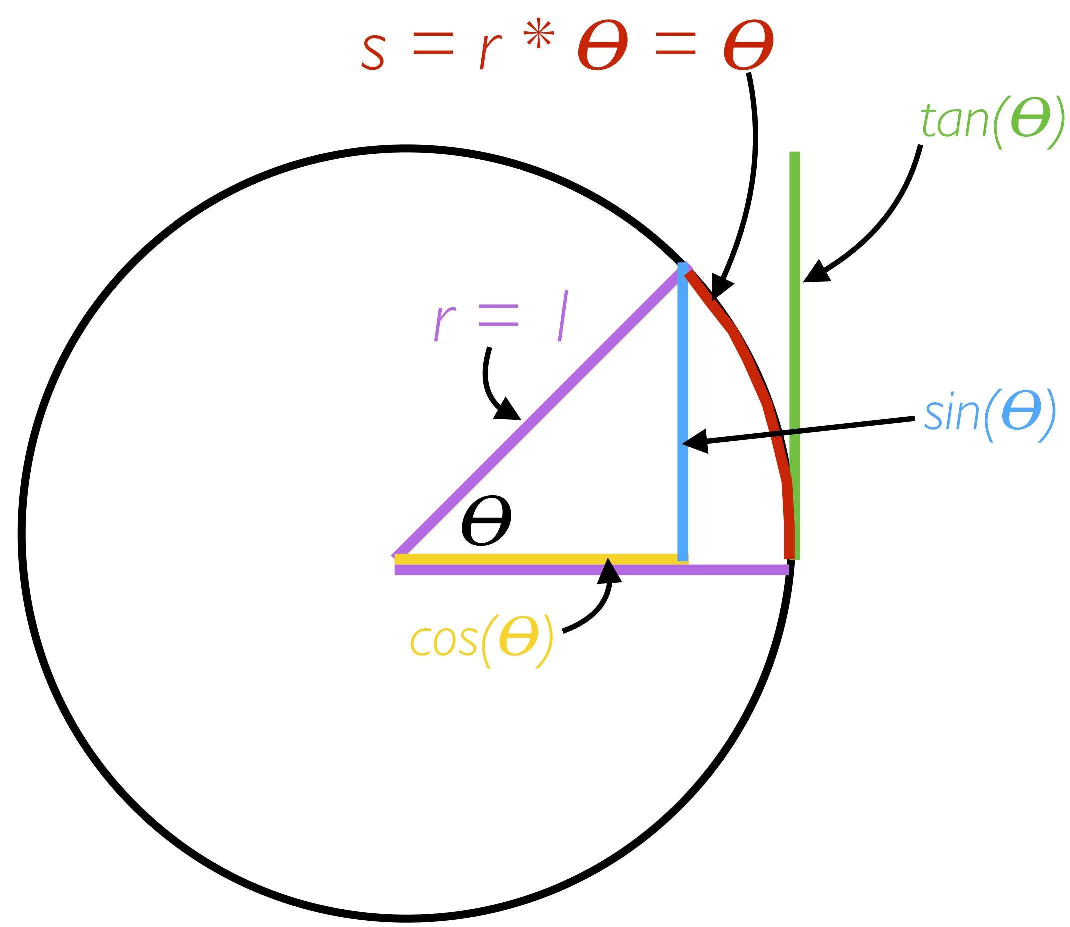



So:

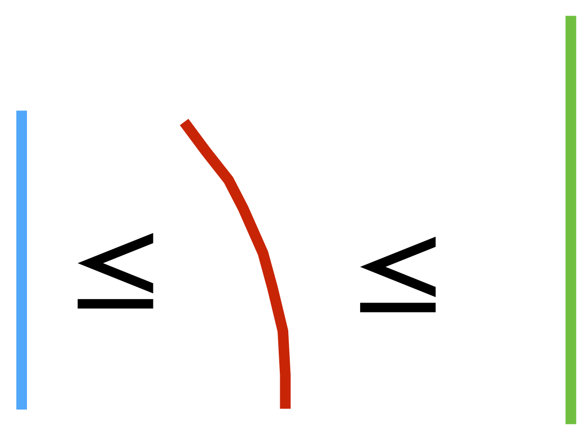
$$\sin(\theta) \leq \theta \leq \tan(\theta)$$

# Taylor Series

Assume circle is radius 1




But obviously, the lengths:




So:

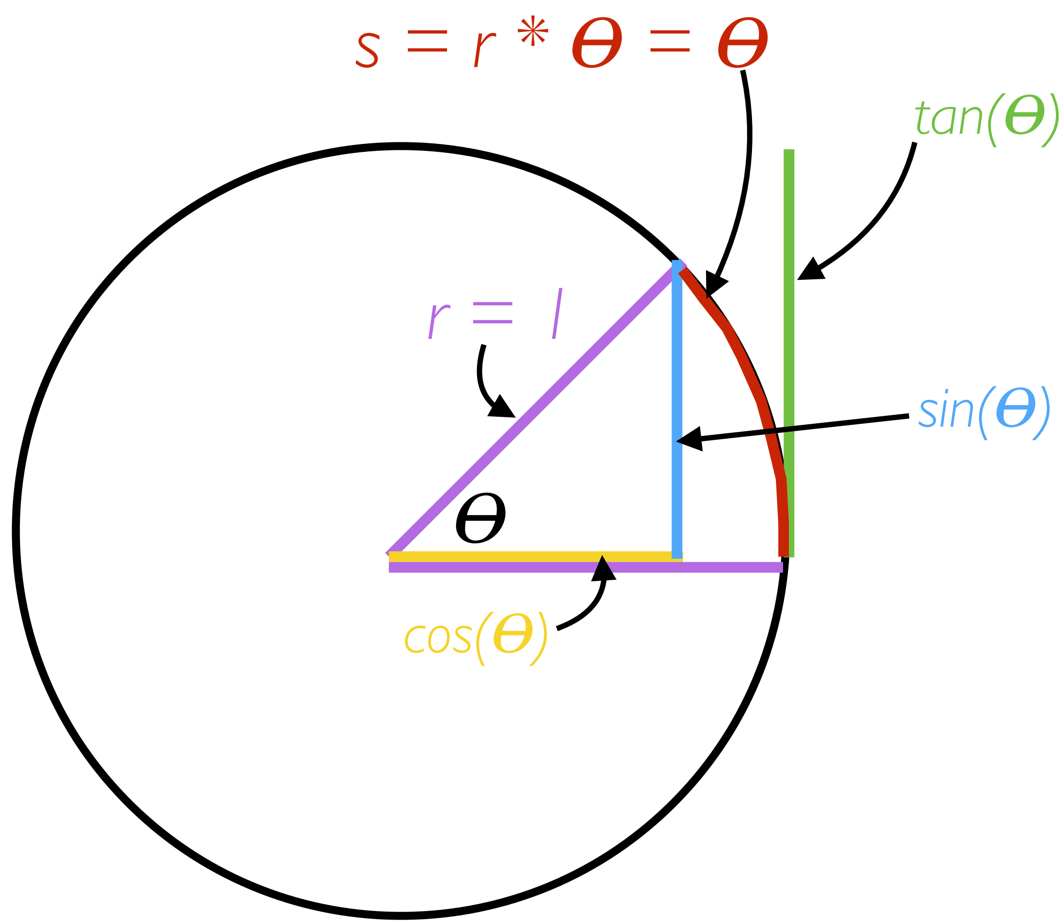
$$\sin(\theta) \leq \theta \leq \tan(\theta)$$

Thus:

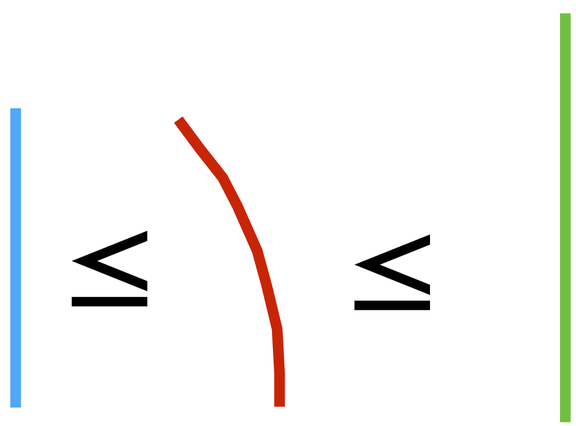
$$1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{\tan(\theta)}{\cos(\theta)} = \frac{1}{\cos(\theta)}$$

# Taylor Series

Assume circle is radius 1




But obviously, the lengths:




So:

$$\sin(\theta) \leq \theta \leq \tan(\theta)$$

Thus:

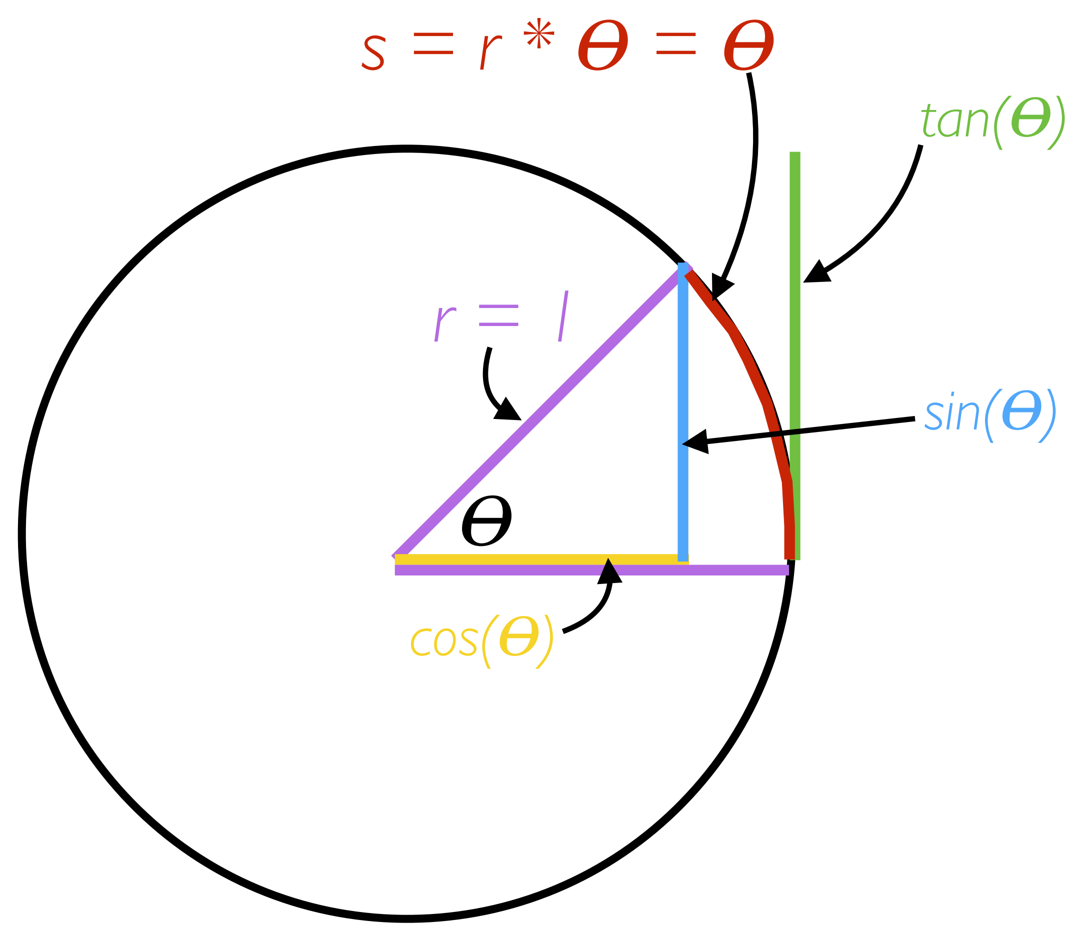
$$1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{\tan(\theta)}{\cos(\theta)} = \frac{1}{\cos(\theta)}$$

When $\theta \to 0$, $\cos(\theta) \to 1$, so:

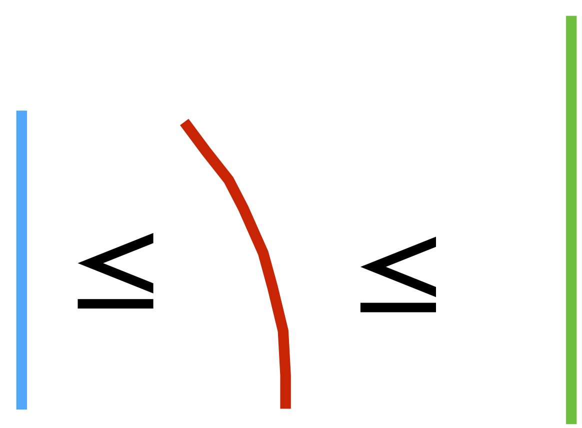
$$1 \leq \lim_{\theta \to 0} \frac{\theta}{\sin(\theta)} \leq 1$$

# Taylor Series

Assume circle is radius 1




But obviously, the lengths:




So:

$$\sin(\theta) \leq \theta \leq \tan(\theta)$$

Thus:

$$1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{\tan(\theta)}{\cos(\theta)} = \frac{1}{\cos(\theta)}$$

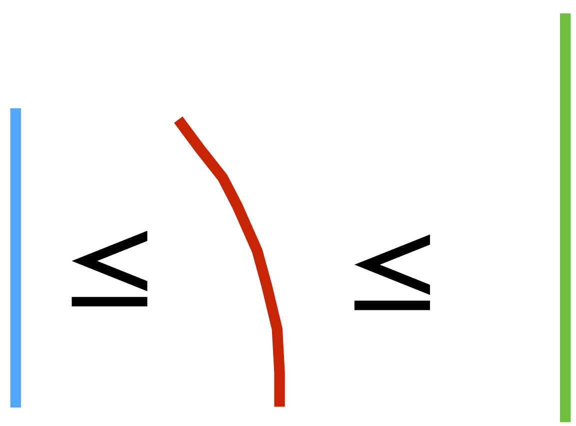
When $\theta \rightarrow 0$, $\cos(\theta) \rightarrow 1$, so:

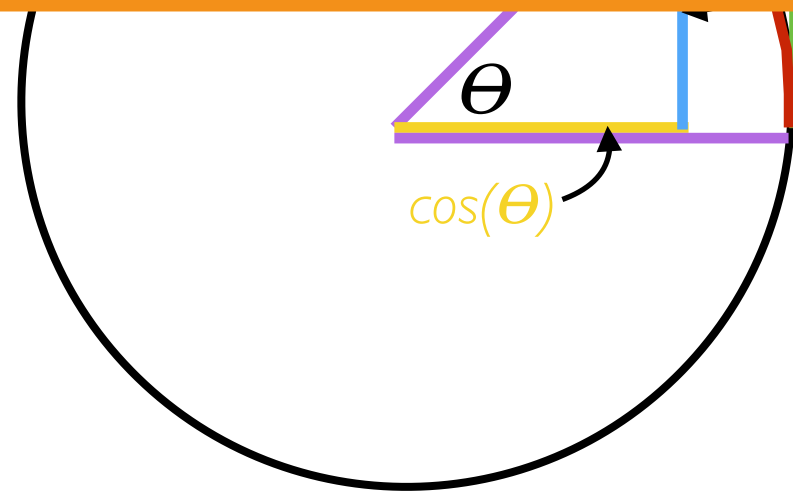
$$1 \leq \lim_{\theta \to 0} \frac{\theta}{\sin(\theta)} \leq 1$$

BEAUTIFUL!!!

# Taylor Series

[Tangential Observation: this calculation of this ***sin(x)/x*** limit much "more specific" mathematically than the ***(cos(x) - 1) / x*** limit, because the latter just depended on ***cos(x)*** being an even function — any even function would have had a similar Taylor series. However, to get ***sin(x)/x*** right, we really needed to understand the specifics of trigonometry.]






So:

$$\sin(\theta) \leq \theta \leq \tan(\theta)$$

Thus:

$$1 \leq \frac{\theta}{\sin(\theta)} \leq \frac{\tan(\theta)}{\cos(\theta)} = \frac{1}{\cos(\theta)}$$

When $\boldsymbol{\theta}$ → 0, cos($\boldsymbol{\theta}$) → 1, so:

$$1 \leq \lim_{\theta \to 0} \frac{\theta}{\sin(\theta)} \leq 1$$

BEAUTIFUL!!!

# Taylor Series

Anyway, we finally have:

$$\sin'(x) = \cos(x) \overbrace{\lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}}^{=\ 1} = \boxed{\cos(x)}$$

# Taylor Series

Anyway, we finally have:

$$\sin'(x) = \cos(x) \overbrace{\lim_{\Delta x \to 0} \frac{\sin(\Delta x)}{\Delta x}}^{= 1} = \boxed{\cos(x)}$$

Similarly we can show:

$$\sin''(x) = \cos'(x) = -\sin(x)$$

Thus:

$$\sin^{(3)}(x) = -\cos(x)$$

And:

$$\sin^{(4)}(x) = \sin(x)$$

# Taylor Series

With that in mind we can calculate full Taylor series' for sine and cosine (around 0):

$$\sin(x) = \sin(0) + \sin'(0)x + \frac{1}{2}\sin^{(2)}(0)x^2 + \ldots$$

# Taylor Series

With that in mind we can calculate full Taylor series' for sine and cosine (around 0):

$$\sin(x) = \sin(0) + \sin'(0)x + \frac{1}{2}\sin^{(2)}(0)x^2 + \ldots$$

$$= 0 + \cos(0)x - \frac{1}{2}\sin(0)x^2 - \frac{1}{3!}\cos(0)x^3 + ...$$

# Taylor Series

With that in mind we can calculate full Taylor series' for sine and cosine (around 0):

$$\begin{aligned}\sin(x) &= \sin(0) + \sin'(0)x + \frac{1}{2}\sin^{(2)}(0)x^2 + \ldots \\ &= 0 + \cos(0)x - \frac{1}{2}\sin(0)x^2 - \frac{1}{3!}\cos(0)x^3 + ... \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots\end{aligned}$$

# Taylor Series

With that in mind we can calculate full Taylor series' for sine and cosine (around 0):

$$\sin(x) = \sin(0) + \sin'(0)x + \frac{1}{2}\sin^{(2)}(0)x^2 + \dots$$

$$= 0 + \cos(0)x - \frac{1}{2}\sin(0)x^2 - \frac{1}{3!}\cos(0)x^3 + ...$$

$$= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots$$

$$= \sum_{i=0}^{\infty} \frac{(-1)^i}{(2i+1)!} x^{2i+1}$$

# Taylor Series

With that in mind we can calculate full Taylor series' for sine and cosine (around 0):

$$\sin(x) = \sin(0) + \sin'(0)x + \frac{1}{2}\sin^{(2)}(0)x^2 + \dots$$

$$= 0 + \cos(0)x - \frac{1}{2}\sin(0)x^2 - \frac{1}{3!}\cos(0)x^3 + \dots$$

$$= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots$$

$$= \sum_{i=0}^{\infty} \frac{(-1)^i}{(2i+1)!} x^{2i+1}$$

And similarly: $$\cos(x) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(2i)!} x^{2i}$$

# Chain Rule

Assume we have a function that's a composition of functions:

$$h(x) = g(f(x))$$

Then the chain rule says:

$$\frac{dh}{dx}(x) = \frac{dg}{dx}(f(x)) \cdot \frac{df}{dx}(x) = g'(f(x)) \cdot f'(x)$$

# Chain Rule

Proof of the Chain Rule:

$$\frac{dh(x)}{dx} = \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x}$$

# Chain Rule

Proof of the Chain Rule:

$$\frac{dh(x)}{dx} = \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(f(x + \Delta x)) - g(f(x))}{\Delta x}$$

# Chain Rule

Proof of the Chain Rule:

$$\frac{dh(x)}{dx} = \lim_{\Delta x \to 0} \frac{h(x+\Delta x) - h(x)}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{g(f(x+\Delta x)) - g(f(x))}{\Delta x}$$

Since by definition:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

we have:

$$f(x+\Delta x) \sim f'(x)\Delta x + f(x)$$

# Chain Rule

Proof of the Chain Rule:

$$\frac{dh(x)}{dx} = \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(f(x + \Delta x)) - g(f(x))}{\Delta x}$$

Since by definition:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

we have:

$$f(x + \Delta x) \sim f'(x)\Delta x + f(x)$$

thus

# Chain Rule

Proof of the Chain Rule:

$$
\begin{aligned}
\frac{dh(x)}{dx} &= \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f(x + \Delta x)) - g(f(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f'(x)\Delta x + f(x)) - g(f(x))}{\Delta x}
\end{aligned}
$$

# Chain Rule

Proof of the Chain Rule:

$$\begin{aligned}\frac{dh(x)}{dx} &= \lim_{\Delta x \to 0} \frac{h(x+\Delta x) - h(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{g(f(x+\Delta x)) - g(f(x))}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{g(f'(x)\Delta x + f(x)) - g(f(x))}{\Delta x}\end{aligned}$$

Now apply the same reasoning to **g** and treating ***f'(x)Δx*** as the "new **Δ*x***" and f(x) as the "new ***x***":

$$g(f'(x)\Delta x + f(x)) \sim g'(f(x))f'(x)\Delta x + g(f(x))$$

# Chain Rule

Proof of the Chain Rule:

$$\frac{dh(x)}{dx} = \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(f(x + \Delta x)) - g(f(x))}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{g(f'(x)\Delta x + f(x)) - g(f(x))}{\Delta x}$$

Now apply the same reasoning to ***g*** and treating ***f'(x)Δx*** as the "new ***Δx***" and f(x) as the "new ***x***":

$$g(f'(x)\Delta x + f(x)) \sim g'(f(x))f'(x)\Delta x + g(f(x))$$

thus

## Chain Rule

Proof of the Chain Rule:

$$
\begin{aligned}
\frac{dh(x)}{dx} &= \lim_{\Delta x \to 0} \frac{h(x+\Delta x) - h(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f(x+\Delta x)) - g(f(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f'(x)\Delta x + f(x)) - g(f(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g'(f(x))f'(x)\Delta x + g(f(x)) - g(f(x))}{\Delta x}
\end{aligned}
$$

## Chain Rule

Proof of the Chain Rule:

$$\begin{aligned}
\frac{dh(x)}{dx} &= \lim_{\Delta x \to 0} \frac{h(x + \Delta x) - h(x)}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f(x + \Delta x)) - g(f(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g(f'(x)\Delta x + f(x)) - g(f(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{g'(f(x))f'(x)\Delta x + g(f(x)) - g(f(x))}{\Delta x} \\
&= g'(f(x))f'(x)
\end{aligned}$$

# Chain Rule

But wait, what exactly did I mean by:

$$f(x + \Delta x) \sim f'(x)\Delta x + f(x)$$

# Chain Rule

But wait, what exactly did I mean by:

$$f(x+\Delta x) \sim f'(x)\Delta x + f(x)$$

The right way to interpret this is actually in terms of Taylor Series. Really, I meant:

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + (\Delta x)^2 \cdot (\text{something})$$

# Chain Rule

But wait, what exactly did I mean by:

$$f(x + \Delta x) \sim f'(x)\Delta x + f(x)$$

The right way to interpret this is actually in terms of Taylor Series. Really, I meant:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + (\Delta x)^2 \cdot (\text{something})$$

where the "something" doesn't blow up to infinity when ***Δx → 0***.

# Chain Rule

But wait, what exactly did I mean by:

$$f(x + \Delta x) \sim f'(x)\Delta x + f(x)$$

The right way to interpret this is actually in terms of Taylor Series. Really, I meant:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + (\Delta x)^2 \cdot (\text{something})$$

where the "something" doesn't blow up to infinity when $\boldsymbol{\Delta x \to 0}$.

In other words, ***f(x) + f'(x)Δx*** is the "approximation of ***f(x+ Δx)*** to first order in ***Δx***."

# Chain Rule

But wait, what exactly did I mean by:

$$f(x + \Delta x) \sim f'(x)\Delta x + f(x)$$

The right way to interpret this is actually in terms of Taylor Series. Really, I meant:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + (\Delta x)^2 \cdot (\text{something})$$

where the "something" doesn't blow up to infinity when ***Δx → 0***.

In other words, ***f(x) + f'(x)Δx*** is the "approximation of ***f(x+ Δx)*** to first order in ***Δx***."

Really, the proof of the Chain Rule is essentially a nested set of two Taylor Series approximations, one for ***f*** and the other for ***g***.

# Multi-variable Functions

Let ***F(x, y)*** be a real-valued function of two real variables ***x*** and ***y***.

# Multi-variable Functions

Let ***F(x, y)*** be a real-valued function of two real variables ***x*** and ***y***. By def'n:

$$F_x = \frac{\partial F(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x, y) - F(x,y)}{\Delta x}$$

$$F_y = \frac{\partial F(x,y)}{\partial y} = \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x,y)}{\Delta y}$$

# Multi-variable Functions

Let ***F(x, y)*** be a real-valued function of two real variables ***x*** and **y**. By def'n:

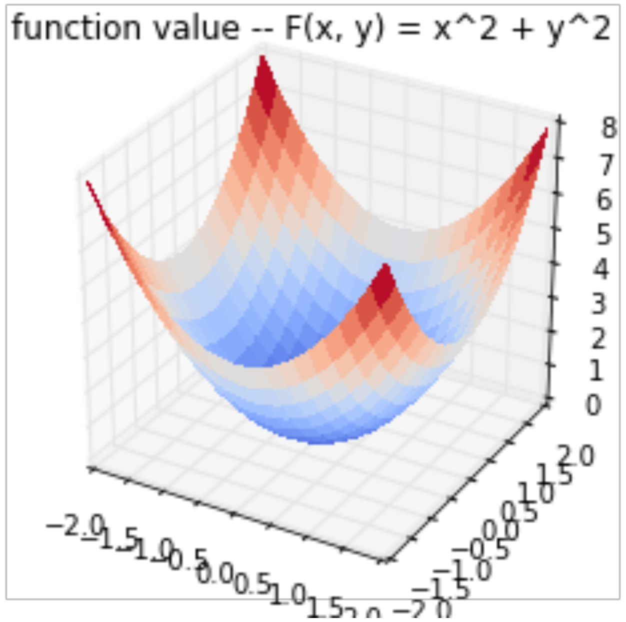
$$F_x = \frac{\partial F(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x, y) - F(x,y)}{\Delta x}$$

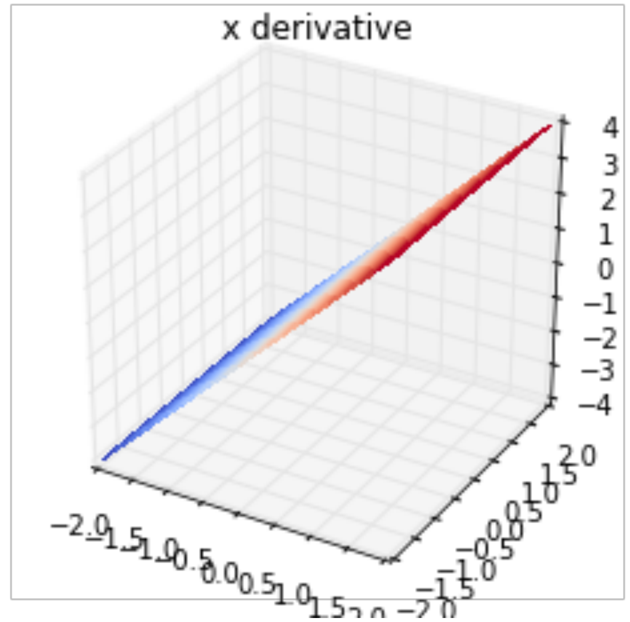
$$F_y = \frac{\partial F(x,y)}{\partial y} = \lim_{\Delta y \to 0} \frac{F(x, y + \Delta y) - F(x,y)}{\Delta y}$$

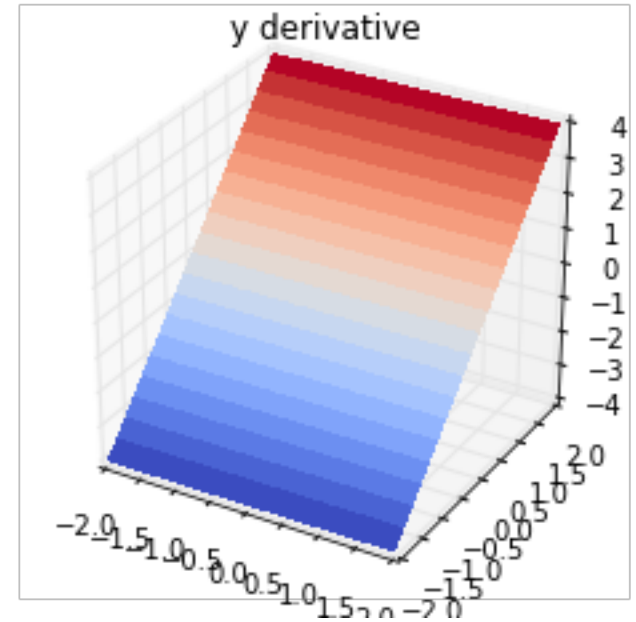
Partial derivative is w.r.t. ***x*** is just derivative of the function holding **y** constant

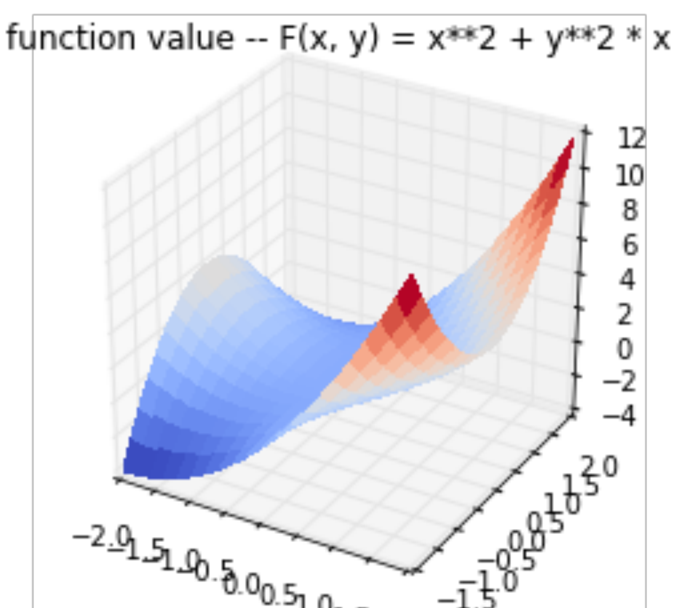
… and vice versa for partial derivative w.r.t. **y**

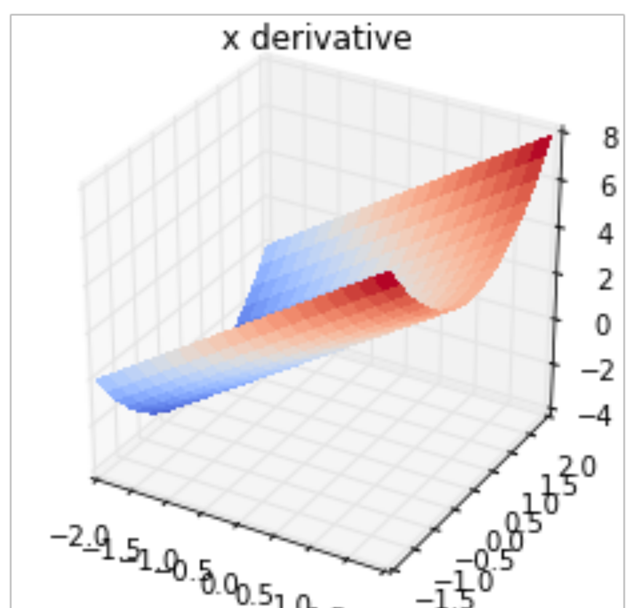
# Multi-variable Functions

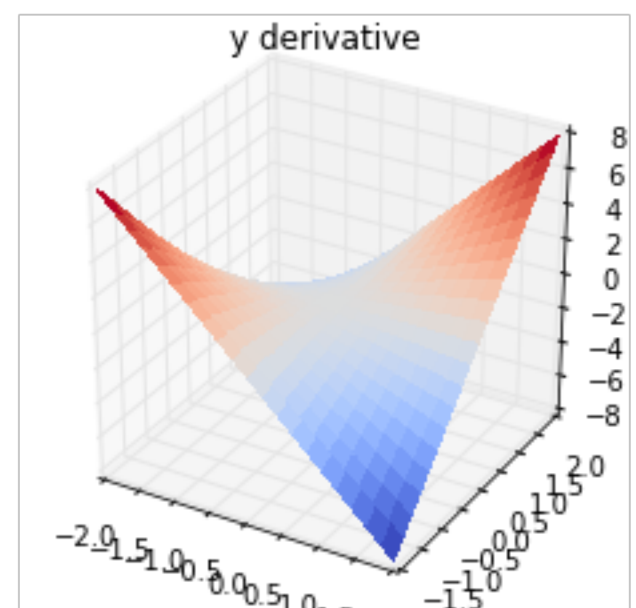


We can also calculate second partial derivatives:

$$F_{xx} = \frac{\partial^2 F(x,y)}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x}\right)(x,y)$$

$$F_{xy} = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right)(x,y)$$

$$F_{yx} = \frac{\partial^2 F(x,y)}{\partial y \partial x} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right)(x,y)$$

$$F_{yy} = \frac{\partial^2 F(x,y)}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y}\right)(x,y)$$

# Multi-variable Functions

We can also calculate second partial derivatives:

$$F_{xx} = \frac{\partial^2 F(x,y)}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x}\right)(x,y)$$

$$F_{xy} = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x}\right)(x,y)$$

$$\parallel$$

$$F_{yx} = \frac{\partial^2 F(x,y)}{\partial y \partial x} = \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right)(x,y)$$

$$F_{yy} = \frac{\partial^2 F(x,y)}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y}\right)(x,y)$$

# Multi-variable Functions

Let ***F(x1, x2, ..., xn)*** be a real-valued function of **n** variables.

The "gradient of ***F***" is the ***n***-vector-valued function:

$$\nabla F(\vec{x}) = \left[\frac{\partial F}{\partial x_1}(\vec{x}), \ldots, \frac{\partial F}{\partial x_n}(\vec{x})\right]$$

# Multi-variable Functions

Let ***F(x1, x2, …, xn)*** be a real-valued function of **n** variables.

The "gradient of ***F***" is the ***n***-vector-valued function:

$$\nabla F(\vec{x}) = \left[\frac{\partial F}{\partial x_1}(\vec{x}), \ldots, \frac{\partial F}{\partial x_n}(\vec{x})\right]$$

the "hessian" matrix

$$H[f] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

shape = (***n***, ***n***)

# Multi-variable Functions

function itself ("0th"-derivative):

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}$$

gradient (1st derivative):

$$\nabla F(\vec{x}) = \left[\frac{\partial F}{\partial x_1}(\vec{x}), \ldots, \frac{\partial F}{\partial x_n}(\vec{x})\right] \qquad \nabla F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

shape = (***n***,)

hessian (2nd derivative):

$$H[f] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \qquad H[F] : \mathbb{R}^n \longrightarrow \mathbb{R}^{n \times n}$$

shape = (***n***, ***n***)

# Multi-variable Functions

function itself ("0th"-derivative):

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}$$

can be generalized to k-tensor-valued function for k-th derivative ... but third derivatives and higher essentially never used...

gradient (1st derivative):

$$\nabla F(\vec{x}) = \left[\frac{\partial F}{\partial x_1}(\vec{x}), \ldots, \frac{\partial F}{\partial x_n}(\vec{x})\right] \qquad \nabla F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

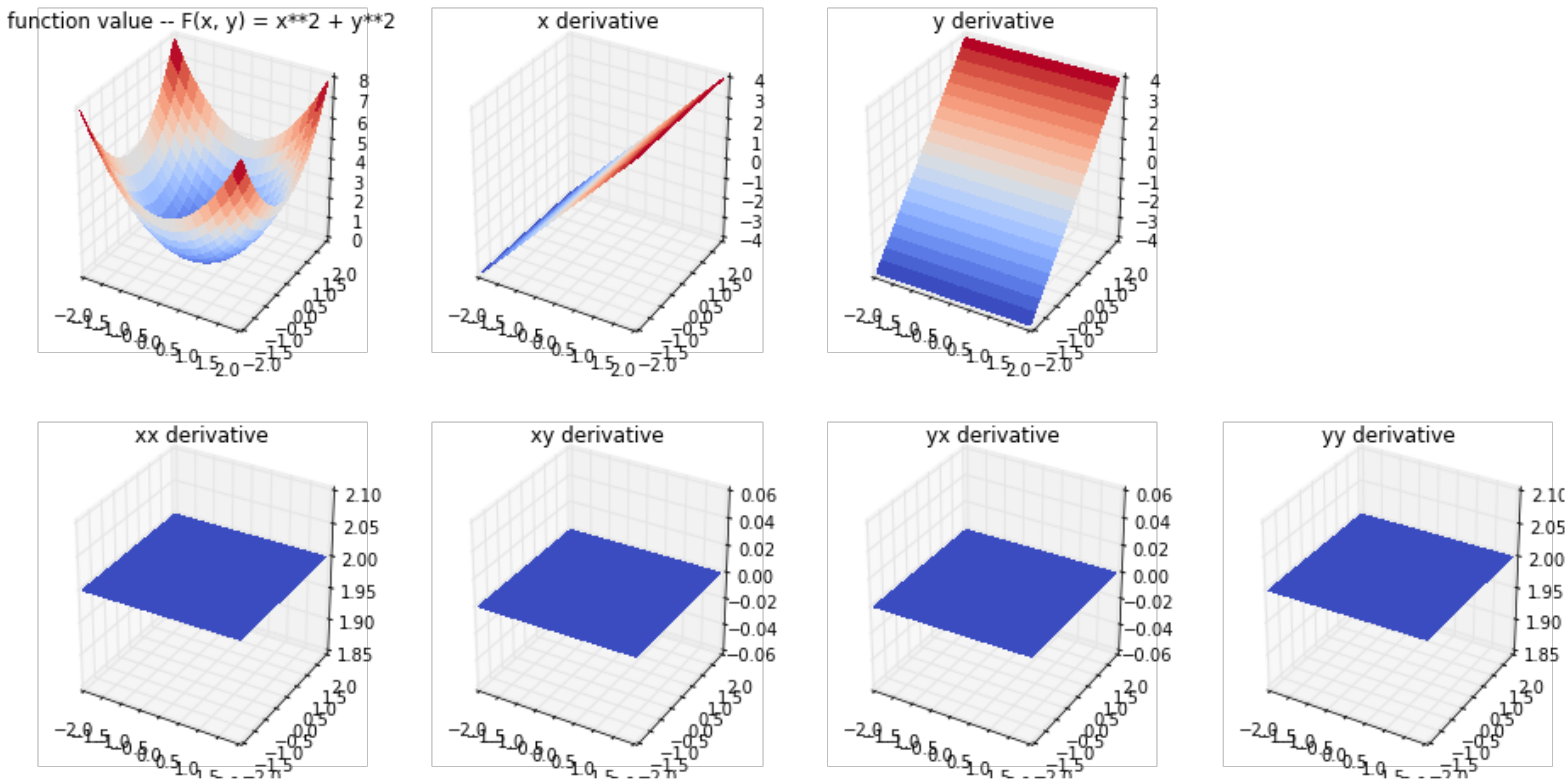
shape = (***n***,)

hessian (2nd derivative):

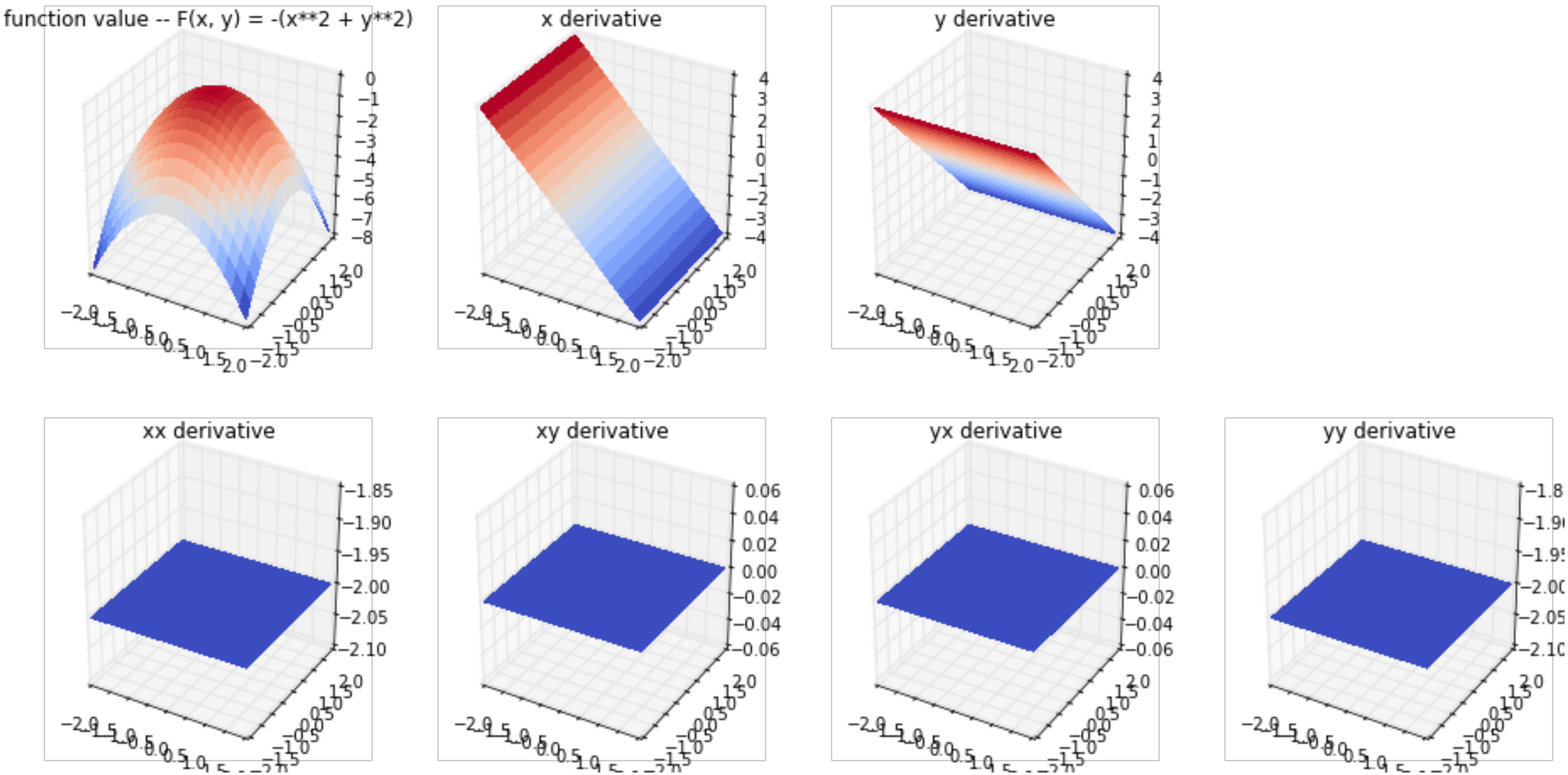
$$H[f] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \qquad H[F] : \mathbb{R}^n \longrightarrow \mathbb{R}^{n \times n}$$

shape = (***n***, ***n***)

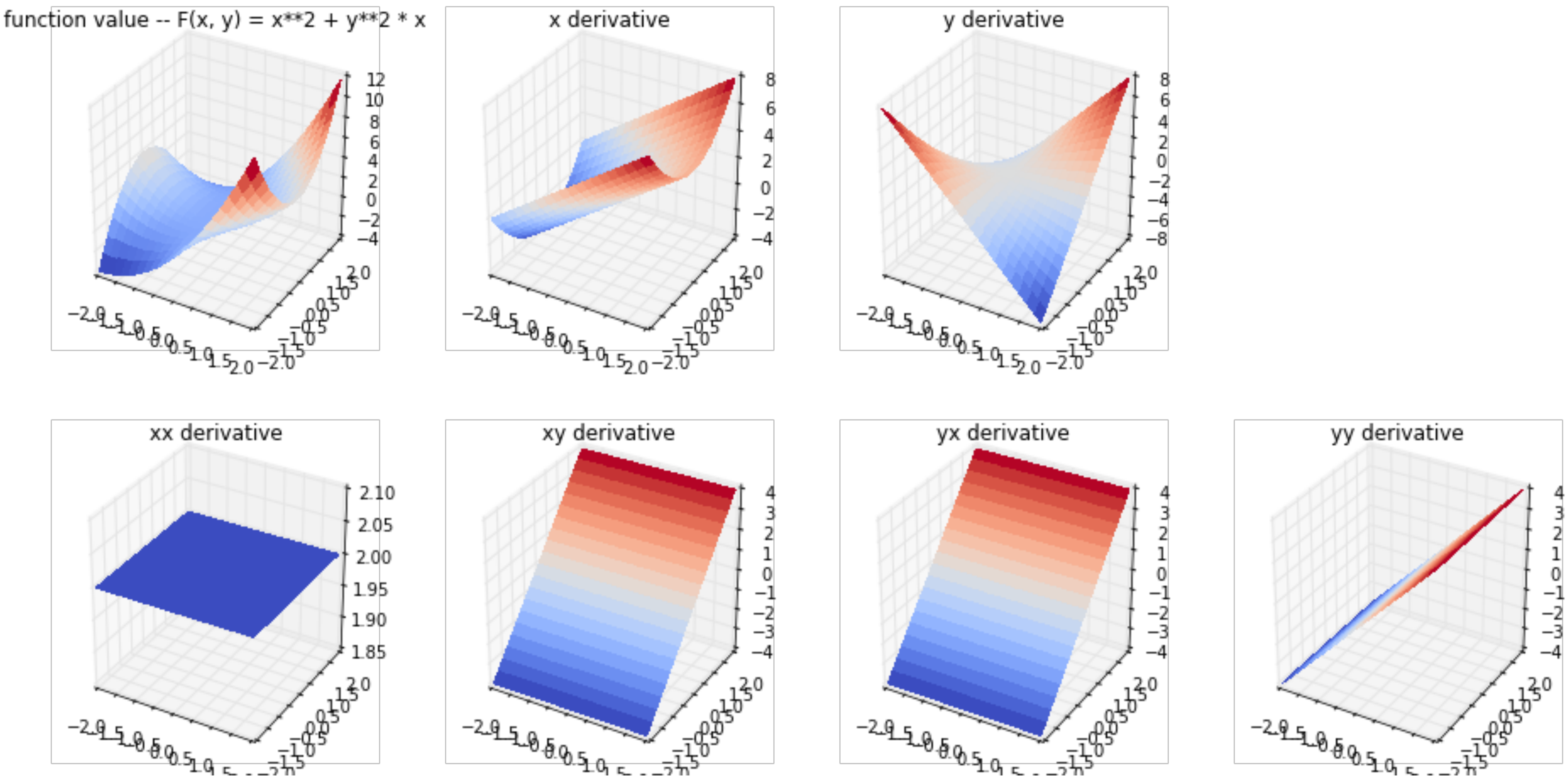
# Multi-variable Functions



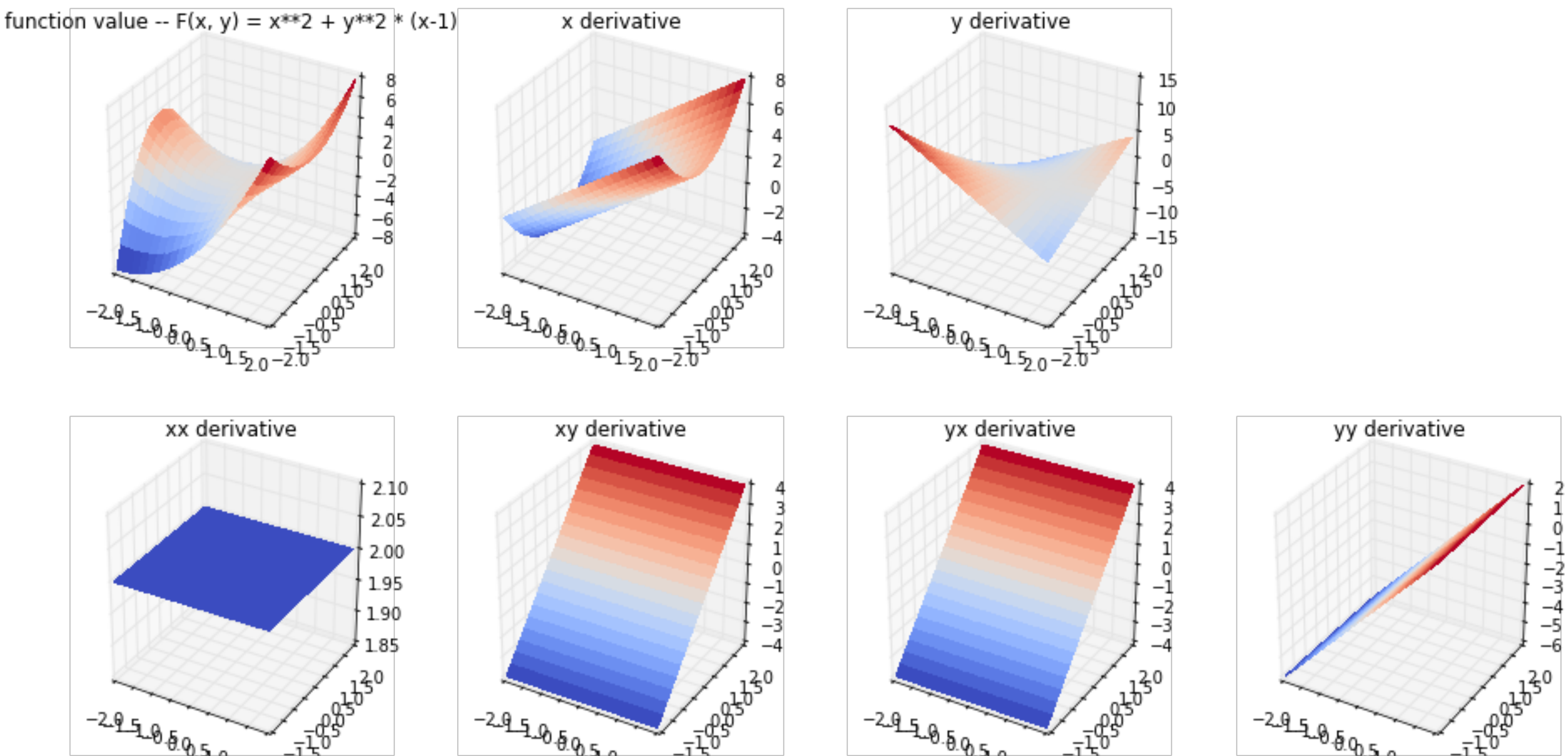
# Multi-variable Functions



# Multi-variable Functions



# Multi-variable Functions



# Multi-variable Functions

In one dimension, critical points are those ***x**** such that:

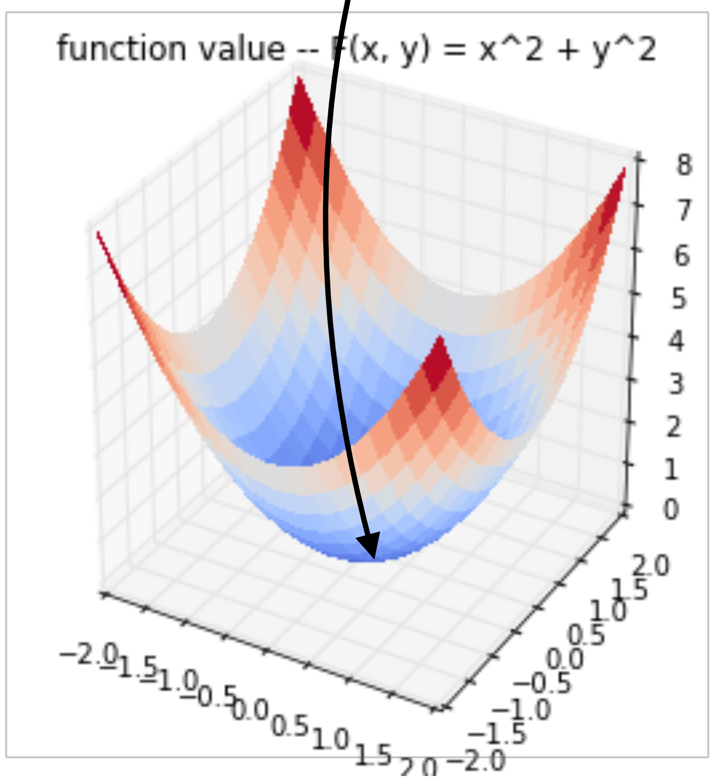
$$\left.\frac{dF}{dx}\right|_{x=x^*} = 0$$

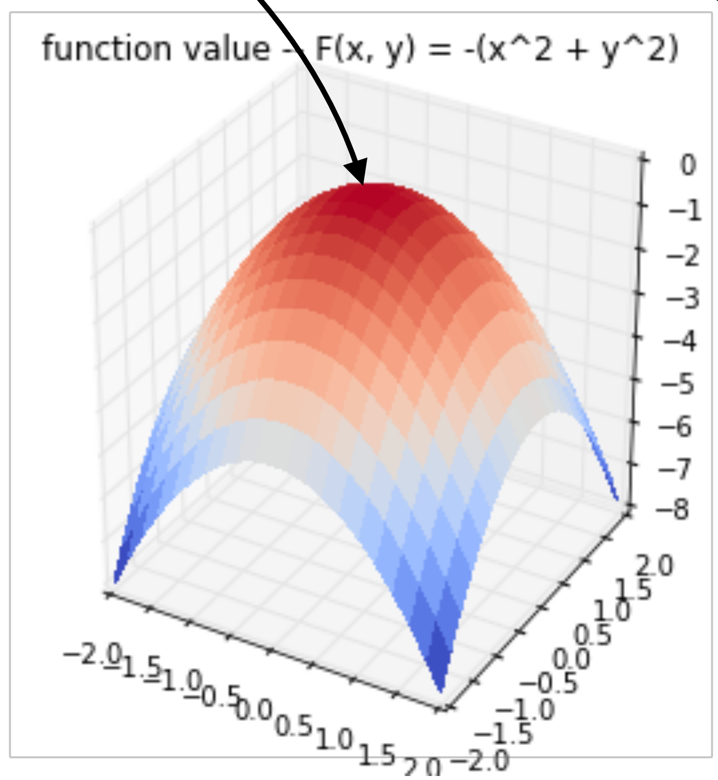
In multiple dimensions, critical points are those ***x**** such that:

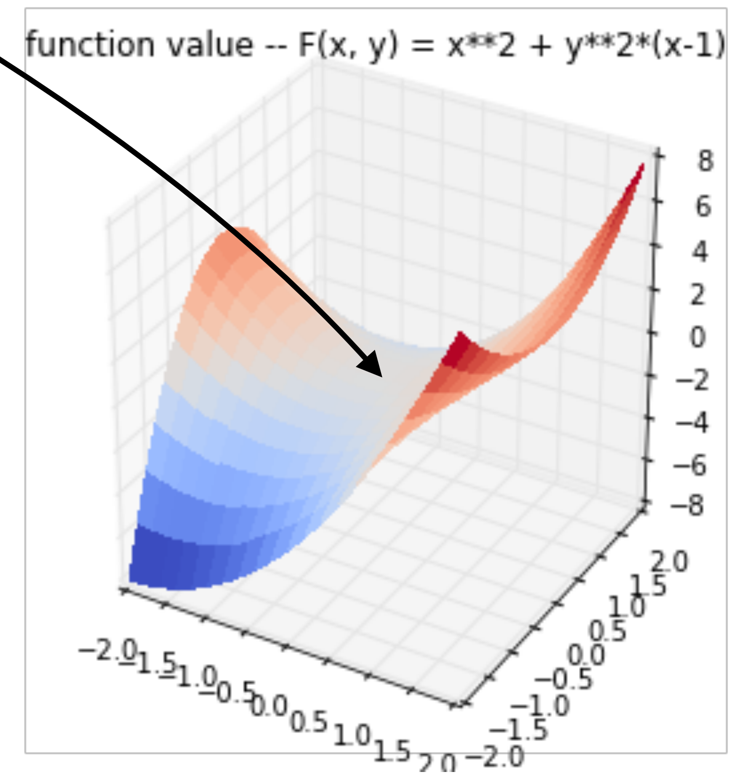
$$\nabla F(\vec{x}^*) = \left[\frac{\partial F}{\partial x_1}(\vec{x}^*), \ldots, \frac{\partial F}{\partial x_n}(\vec{x}^*)\right] = [0, \ldots, 0] = \vec{0}$$

# Multi-variable Functions

Critical point



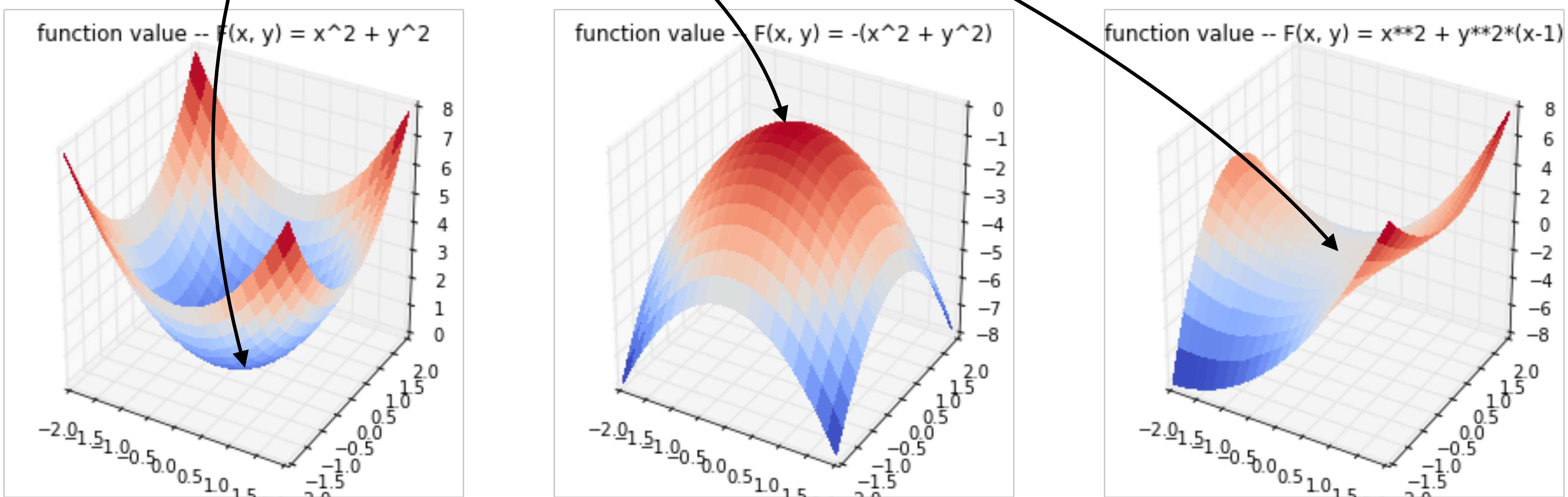



# Multi-variable Functions

Critical point

But what's the difference?




Hessian determinant positive

Hessian determinant positive

Hessian determinant negative

# Multi-variable Functions

Imagine you have a path through variable space as a function of a single variable (like time):

$$\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$$

# Multi-variable Functions

Imagine you have a path through variable space as a function of a single variable (like time):

$$\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$$

Then

$$f(t) = \vec{F}(\vec{x}(t))$$

is a one-variable real-valued function.

# Multi-variable Functions

Imagine you have a path through variable space as a function of a single variable (like time):

$$\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$$

Then

$$f(t) = \vec{F}(\vec{x}(t))$$

is a one-variable real-valued function.

What's its derivative?

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t + \Delta t), \ldots, x_n(t + \Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* ***Δt***

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* ***Δt***

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*taking the limit in the first term*

$$= \frac{\partial F}{\partial x_1}(x_1(t), x_2(t), \ldots, x_n(t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* ***Δt***

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*taking the limit in the first term*

$$= \frac{\partial F}{\partial x_1}(x_1(t), x_2(t), \ldots, x_n(t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*applying same reasoning to* **x2**

$$= \frac{\partial F}{\partial x_1}(\vec{x}(t))x_1'(t) + \frac{\partial F}{\partial x_2}(\vec{x}(t))x_2'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t), x_3(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* ***Δt***

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*taking the limit in the first term*

$$= \frac{\partial F}{\partial x_1}(x_1(t), x_2(t), \ldots, x_n(t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*applying same reasoning to* **x2**

$$= \frac{\partial F}{\partial x_1}(\vec{x}(t))x_1'(t) + \frac{\partial F}{\partial x_2}(\vec{x}(t))x_2'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t), x_3(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*… then to all terms*

*= 0 by def'n of f(t)*

$$= \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) + \cancel{\lim_{\Delta t \to 0} \frac{F(\vec{x}(t)) - f(t)}{\Delta t}} = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) = \nabla F \cdot \vec{x}'(t)$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* $\mathbf{\Delta t}$

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*taking the limit in the first term*

$$= \frac{\partial F}{\partial x_1}(x_1(t), x_2(t), \ldots, x_n(t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*applying same reasoning to* **x2**

$$= \frac{\partial F}{\partial x_1}(\vec{x}(t))x_1'(t) + \frac{\partial F}{\partial x_2}(\vec{x}(t))x_2'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t), x_3(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*… then to all terms*

*= 0 by def'n of f(t)*

$$= \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) + \cancel{\lim_{\Delta t \to 0} \frac{F(\vec{x}(t)) - f(t)}{\Delta t}} = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) = \boxed{\nabla F \cdot \vec{x}'(t)}$$

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\vec{F}(x_1(t+\Delta t), \ldots, x_n(t+\Delta t)) - F(x_1(t), \ldots, x_n(t))}{\Delta t}$$

*applying definitions of derivatives, nested first to F(x) (partial derivative) and then to x(t)*

$$= \lim_{\Delta t \to 0} \frac{\frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t)\Delta t + F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*dividing out* **Δt**

$$= \lim_{\Delta t \to 0} \frac{\partial F}{\partial x_1}(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*taking the limit in the first term*

$$= \frac{\partial F}{\partial x_1}(x_1(t), x_2(t), \ldots, x_n(t))x_1'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*applying same reasoning to* **x2**

$$= \frac{\partial F}{\partial x_1}(\vec{x}(t))x_1'(t) + \frac{\partial F}{\partial x_2}(\vec{x}(t))x_2'(t) + \lim_{\Delta t \to 0} \frac{F(x_1(t), x_2(t), x_3(t+\Delta t), \ldots, x_n(t+\Delta t)) - f(t)}{\Delta t}$$

*… then to all terms*

*= 0 by def'n of f(t)*

*dot product of two n vectors, yielding a scalar*

$$= \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) + \cancel{\lim_{\Delta t \to 0} \frac{F(\vec{x}(t)) - f(t)}{\Delta t}} = \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(\vec{x}(t))x_i'(t) = \boxed{\nabla F \cdot \vec{x}'(t)}$$

# Multi-variable Functions

Imagine you have a path through variable space as a function of a single variable (like time):

$$\vec{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t))$$

Then

$$f(t) = \vec{F}(\vec{x}(t))$$

is a one-variable real-valued function. What's its derivative?

Answer:

$$\frac{df}{dt} = \frac{\partial F}{\partial x_1} \cdot \frac{dx_1(t)}{dt} + \ldots + \frac{\partial F}{\partial x_n} \cdot \frac{dx_n(t)}{dt} = \nabla F(x(t)) \cdot \frac{d\vec{x}(t)}{dt}$$

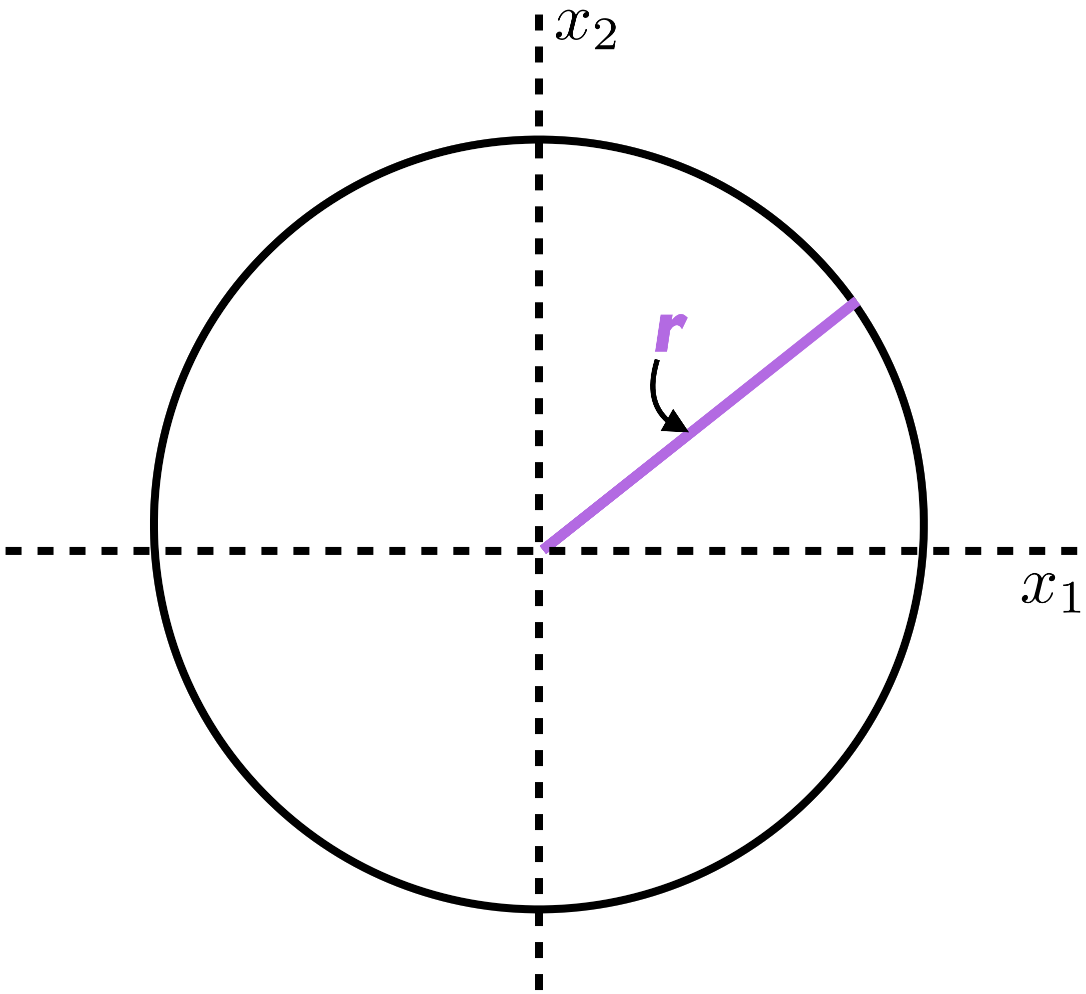
# Multi-variable Functions

With these idea in mind, let's consider the function: $F(x_1, x_2) = x_1^2 + x_2^2$

# Multi-variable Functions

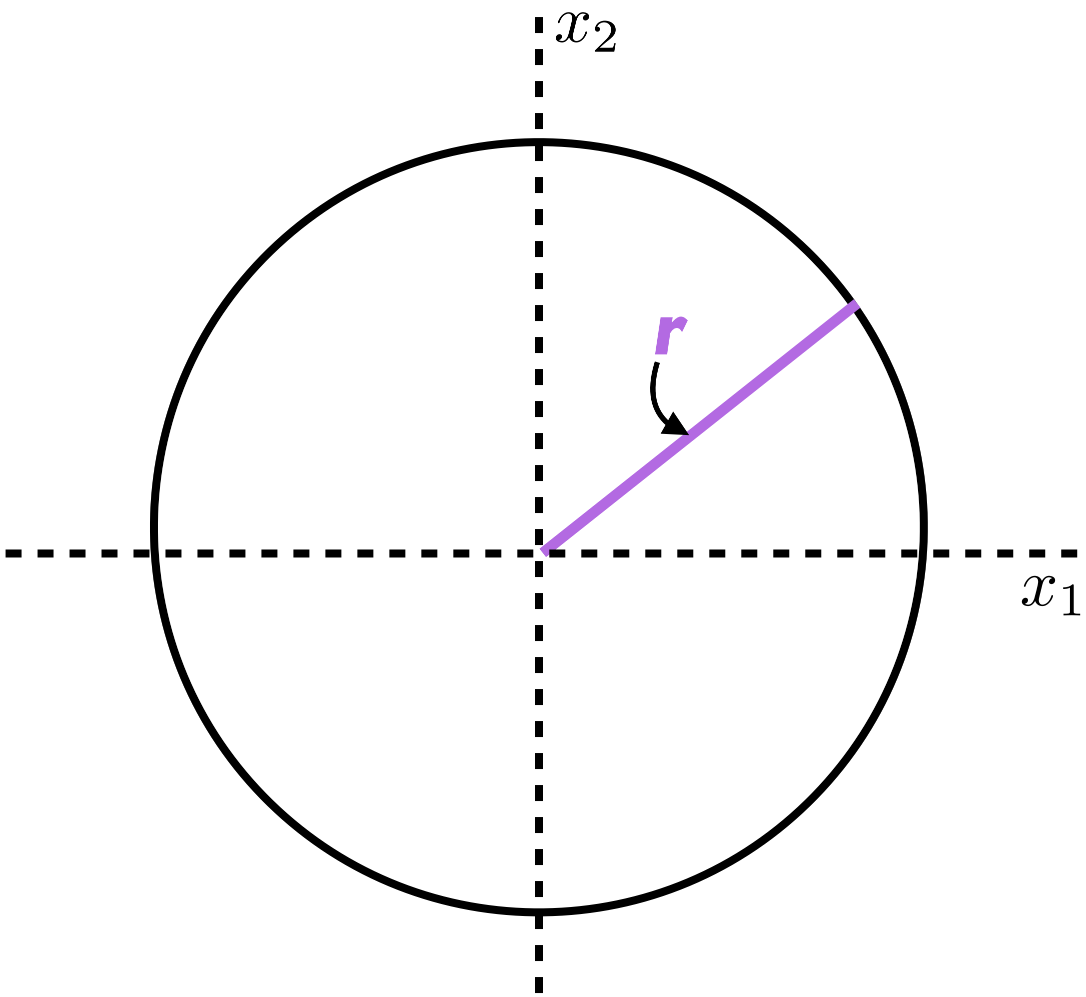
With these idea in mind, let's consider the function: $F(x_1, x_2) = x_1^2 + x_2^2$

The circle radius r is the set of all points where $F(x_1, x_2) = r^2$



With these idea in mind, let's consider the function: $F(x_1, x_2) = x_1^2 + x_2^2$

The circle radius r is the set of all points where $F(x_1, x_2) = r^2$



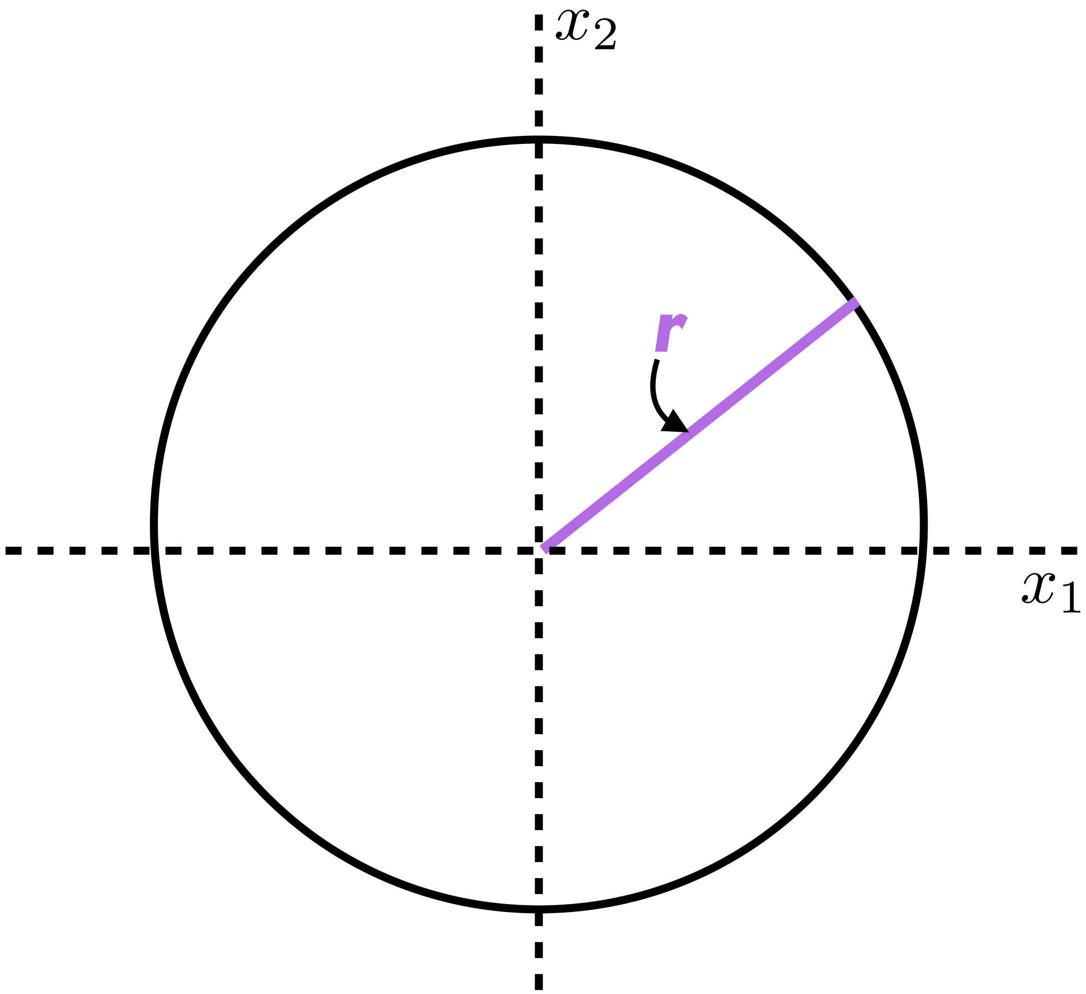

Now consider the path:

$$x(t) = r\cos(t); y(t) = r\sin(t)$$

# Multi-variable Functions

With these idea in mind, let's consider the function: $F(x_1, x_2) = x_1^2 + x_2^2$

The circle radius r is the set of all points where $F(x_1, x_2) = r^2$




Now consider the path:

$$x(t) = r\cos(t); y(t) = r\sin(t)$$

This path is on the circle so ***F*** is constant:

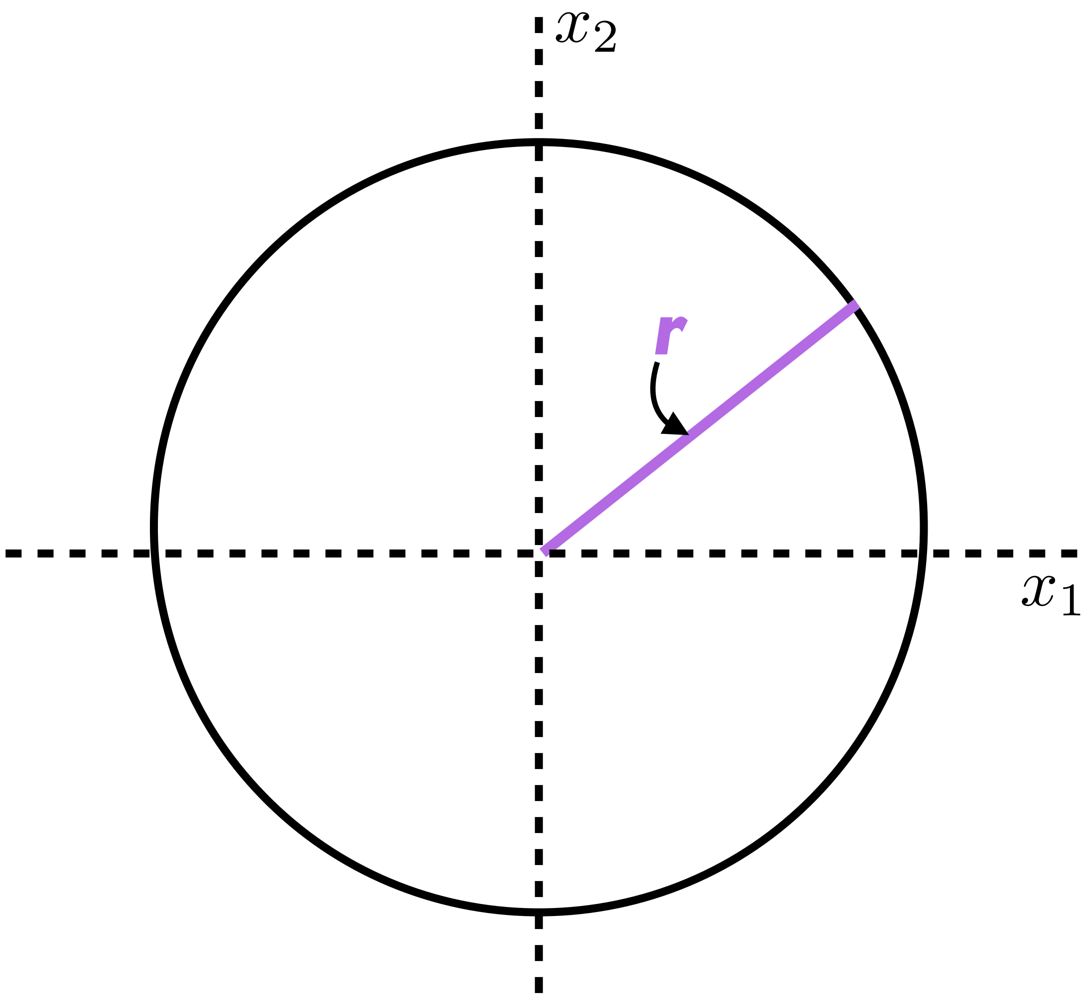
$$F(x(t), y(t)) = [r\cos(t)]^2 + [r\sin(t)]^2$$

$$= r^2(\cos^2(t) + \sin^2(t)) = r^2$$

# Multi-variable Functions

With these idea in mind, let's consider the function: $F(x_1, x_2) = x_1^2 + x_2^2$

The circle radius r is the set of all points where $F(x_1, x_2) = r^2$




Now consider the path:

$$x(t) = r\cos(t); y(t) = r\sin(t)$$

This path is on the circle so ***F*** is constant:

$$F(x(t), y(t)) = [r\cos(t)]^2 + [r\sin(t)]^2$$

$$= r^2(\cos^2(t) + \sin^2(t)) = r^2$$

Thus: $$\frac{df}{dt} = \frac{d}{dt}[F(x(t))] = 0$$ Not a coincidence.

# Multi-variable Functions

In general if a path ***x(t)*** travels along an isocline of ***F***, then:

$$\nabla_x F(x(t)) \cdot \vec{x}'(t) = \frac{d}{dt}[F(x(t)] = 0$$

by the result we proved earlier

since ***F*** is constant along ***x(t)***

# Multi-variable Functions

In general if a path ***x(t)*** travels along an isocline of ***F***, then:

$$\nabla_x F(x(t)) \cdot \vec{x}'(t) = \frac{d}{dt}[F(x(t)] = 0$$

by the result we proved earlier

since ***F*** is constant along ***x(t)***

In other words, the gradient of ***F*** is perpendicular to any path along the isocline:

$$\nabla_x F(x(t)) \perp \vec{x}'(t)$$

# Multi-variable Functions

In general if a path ***x(t)*** travels along an isocline of ***F***, then:

$$\nabla_x F(x(t)) \cdot \vec{x}'(t) = \frac{d}{dt}[F(x(t)] = 0$$

by the result we proved earlier

since ***F*** is constant along ***x(t)***

In other words, the gradient of ***F*** is perpendicular to any path along the isocline:

$$\nabla_x F(x(t)) \perp \vec{x}'(t)$$

But since x(t) was any such path, this means the <u>*gradient vector is always perpendicular to the tangent of the isocline surface*</u>

# Multi-variable Functions

In general if a path ***x(t)*** travels along an isocline of ***F***, then:

$$\nabla_x F(x(t)) \cdot \vec{x}'(t) = \frac{d}{dt}[F(x(t)] = 0$$

by the result we proved earlier

since ***F*** is constant along ***x(t)***

In other words, the gradient of ***F*** is perpendicular to any path along the isocline:

$$\nabla_x F(x(t)) \perp \vec{x}'(t)$$

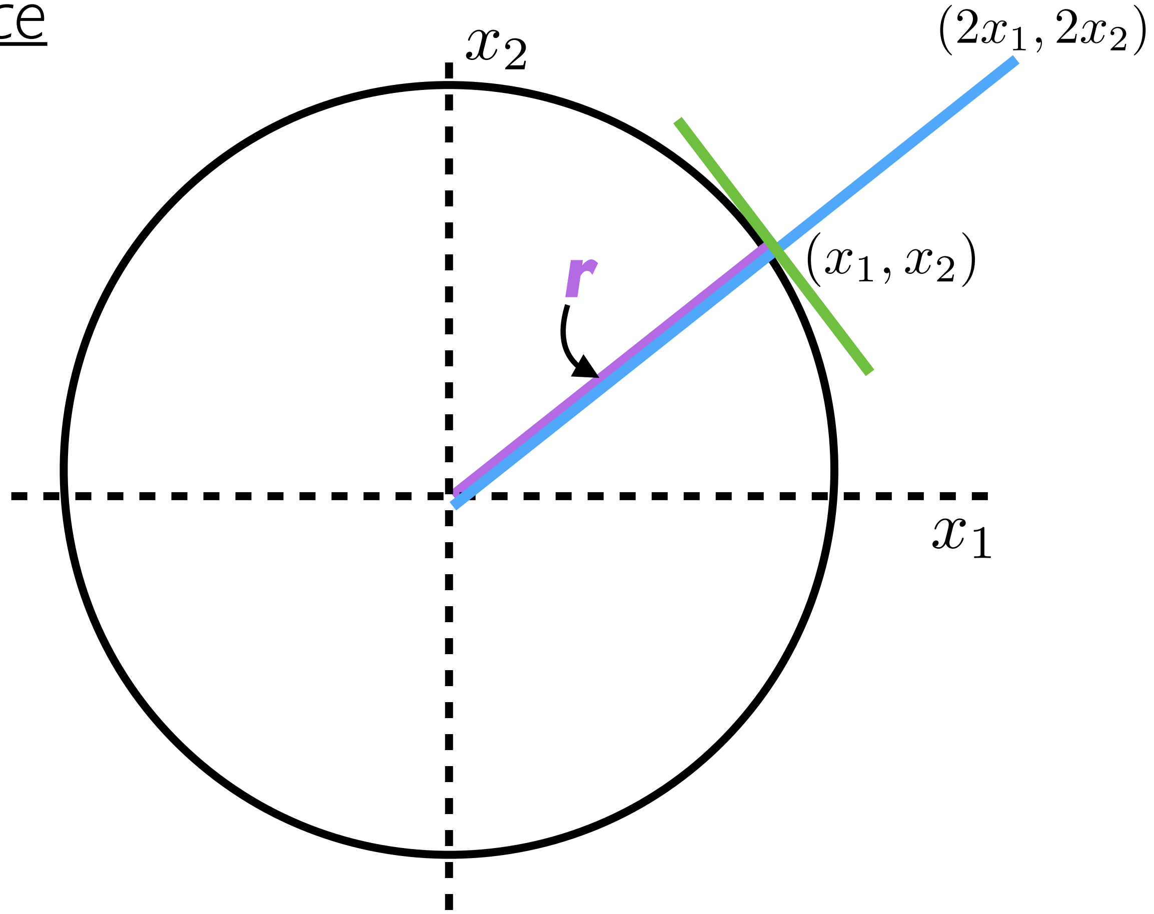
But since x(t) was any such path, this means the ***gradient vector*** *is always perpendicular to the* ***tangent*** *of the isocline surface*

So in our example:

function: $F(x_1, x_2) = x_1^2 + x_2^2$

isocline: $F(x_1, x_2) = r^2$ *(circle of radius r)*

gradient: $\nabla_x F(x_1, x_2) = (2x_1, 2x_2)$



Why do we care about all this? **Constrained Optimization**

# Multi-variable Functions

Why do we care about all this? **Constrained Optimization**

If our goal is to extremize ***F*** we seek critical points x* s.t. $\nabla F(x^*) = \vec{0}$

# Multi-variable Functions

Why do we care about all this? **Constrained Optimization**

If our goal is to extremize ***F*** we seek critical points x* s.t. $\nabla F(x^*) = \vec{0}$

What if our goal is to extremize ***F*** subject to the constraint that ***g = 0***?

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

# Multi-variable Functions

Why do we care about all this? **Constrained Optimization**

If our goal is to extremize ***F*** we seek critical points x* s.t. $\nabla F(x^*) = \vec{0}$

What if our goal is to extremize ***F*** subject to the constraint that ***g = 0***?

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

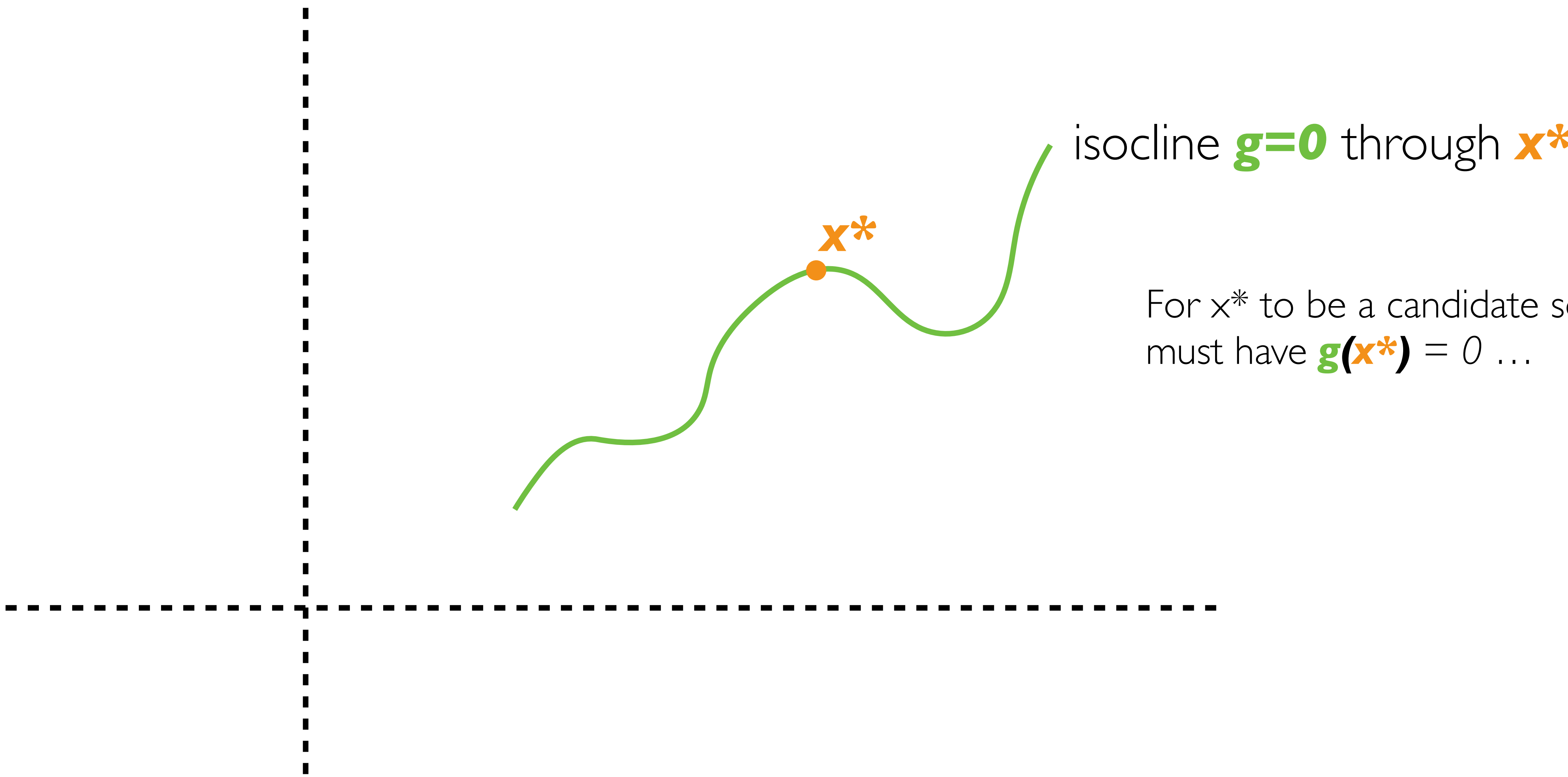
We can't just seek critical points of ***F*** because they might not satisfy the constraint.

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

# Multi-variable Functions

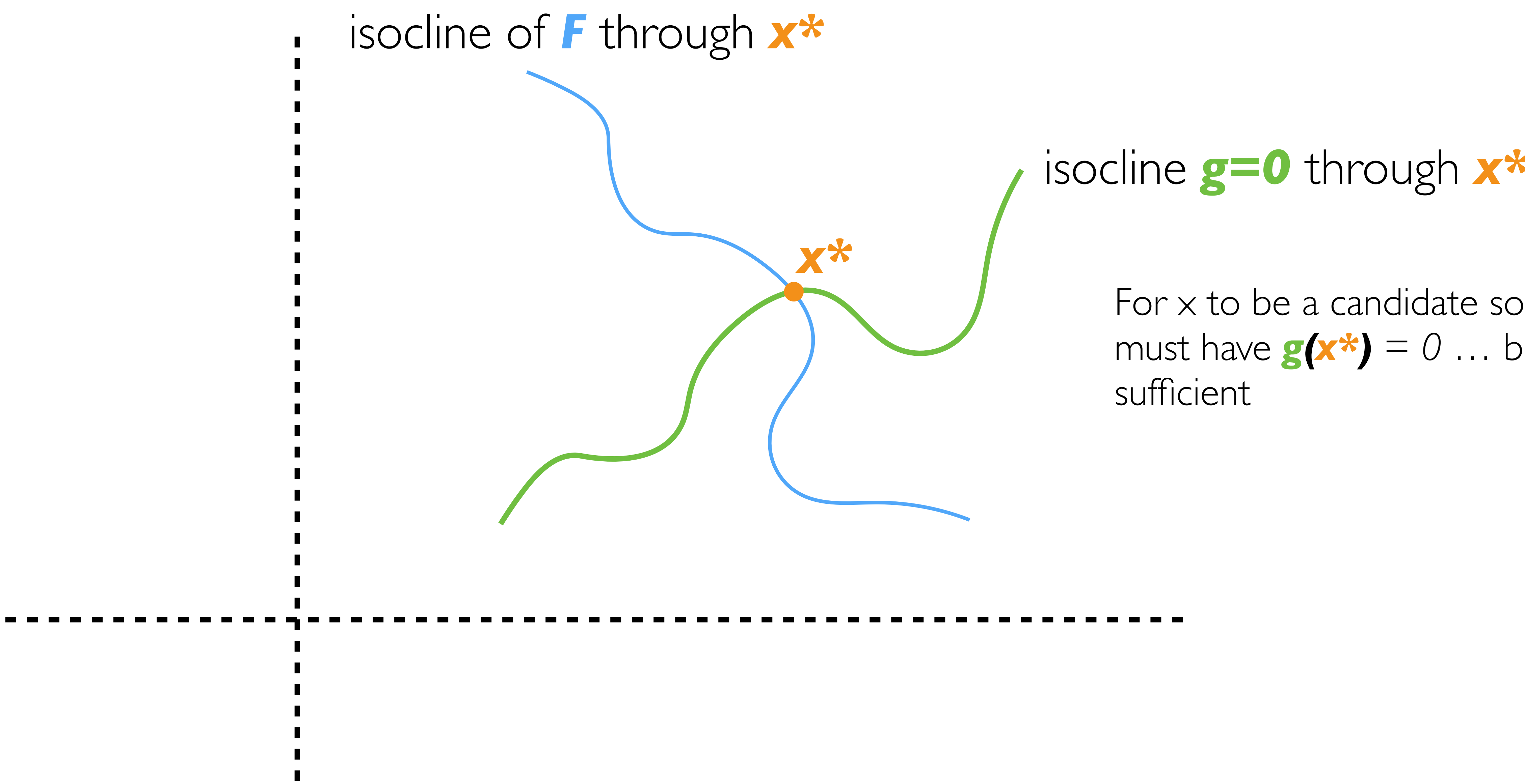
What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?




For x* to be a candidate solution must have $g(x^*) = 0$ …

# Multi-variable Functions

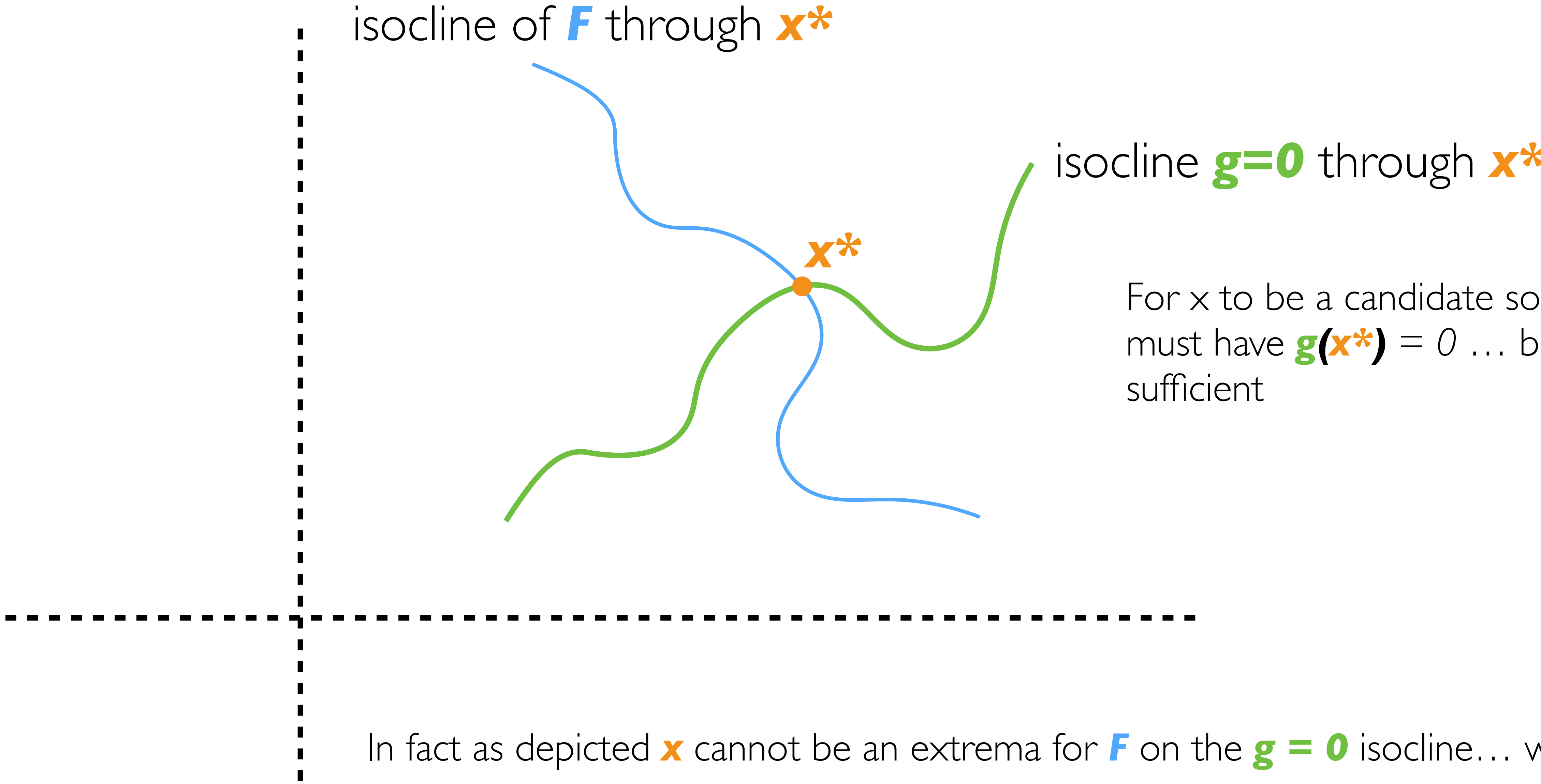
What if our goal is to maximize ***F*** subject to the constraint that **$g = 0$**?




For x to be a candidate solution must have $g(x^*) = 0$ … but not sufficient

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

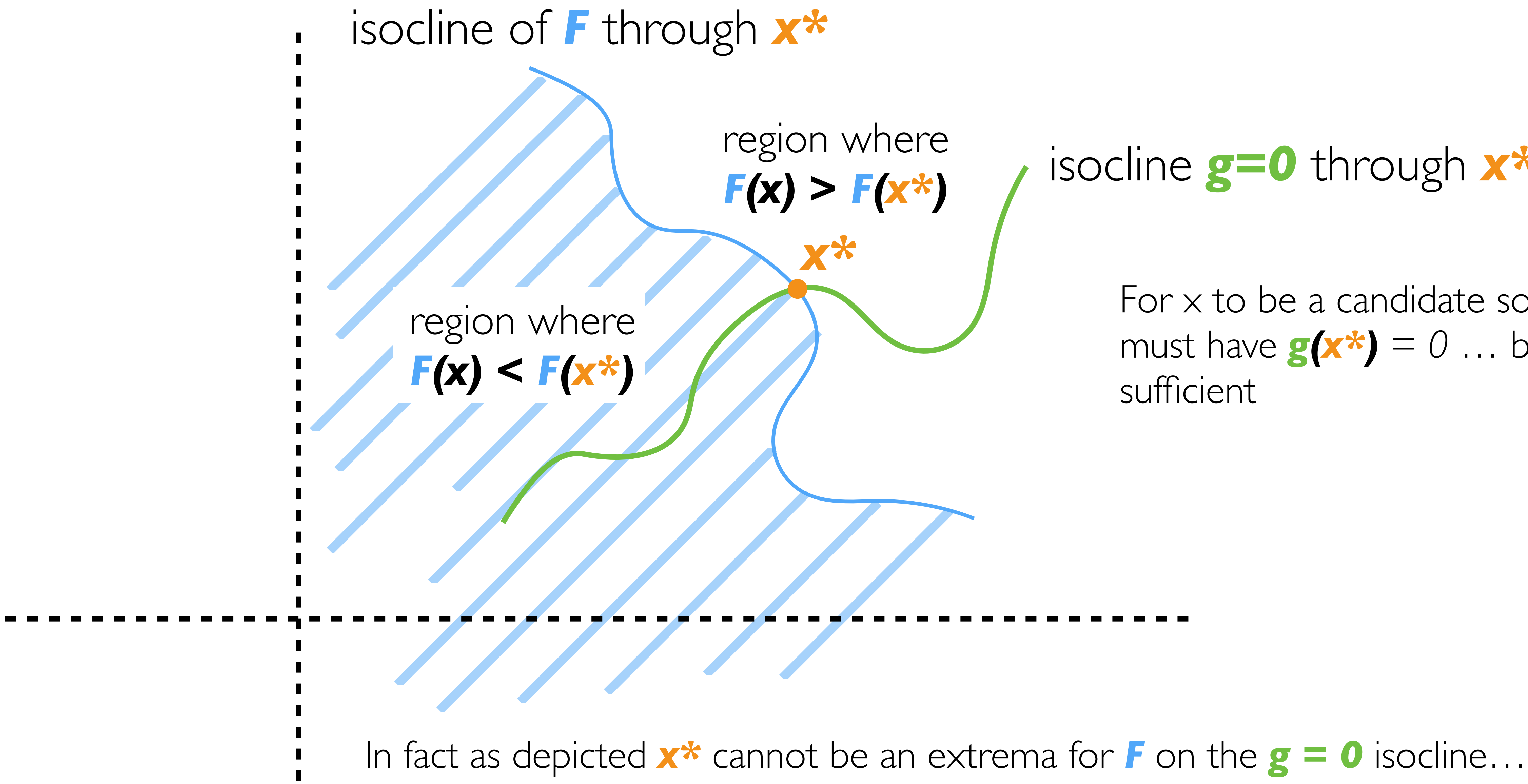



For x to be a candidate solution must have $g(x^*) = 0$ … but not sufficient

In fact as depicted ***x*** cannot be an extrema for ***F*** on the ***g = 0*** isocline… why?

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g* = *0***?

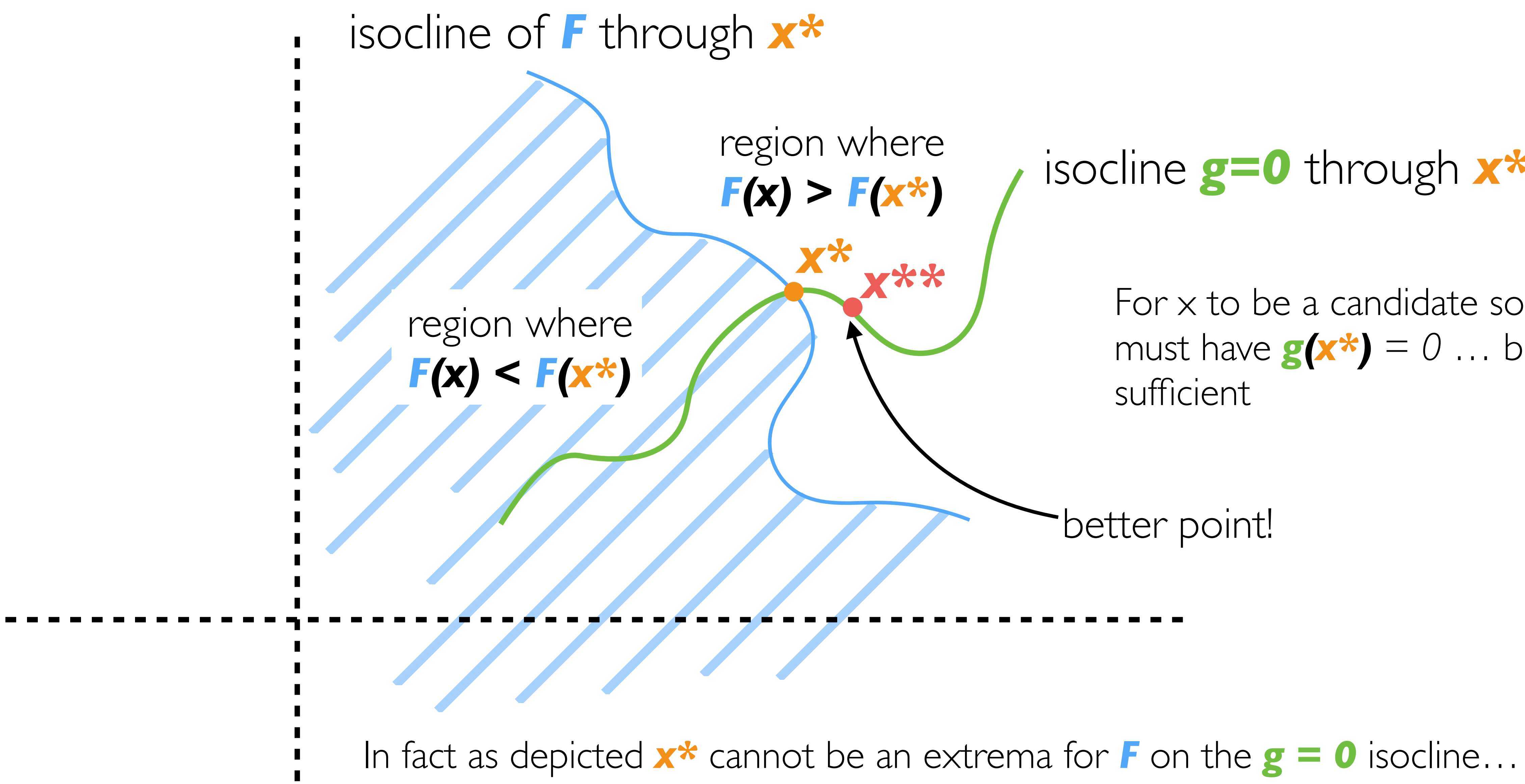



For x to be a candidate solution must have ***g*(*x**)** = *0* … but not sufficient

In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g* = *0*** isocline… why?

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g* = *0***?



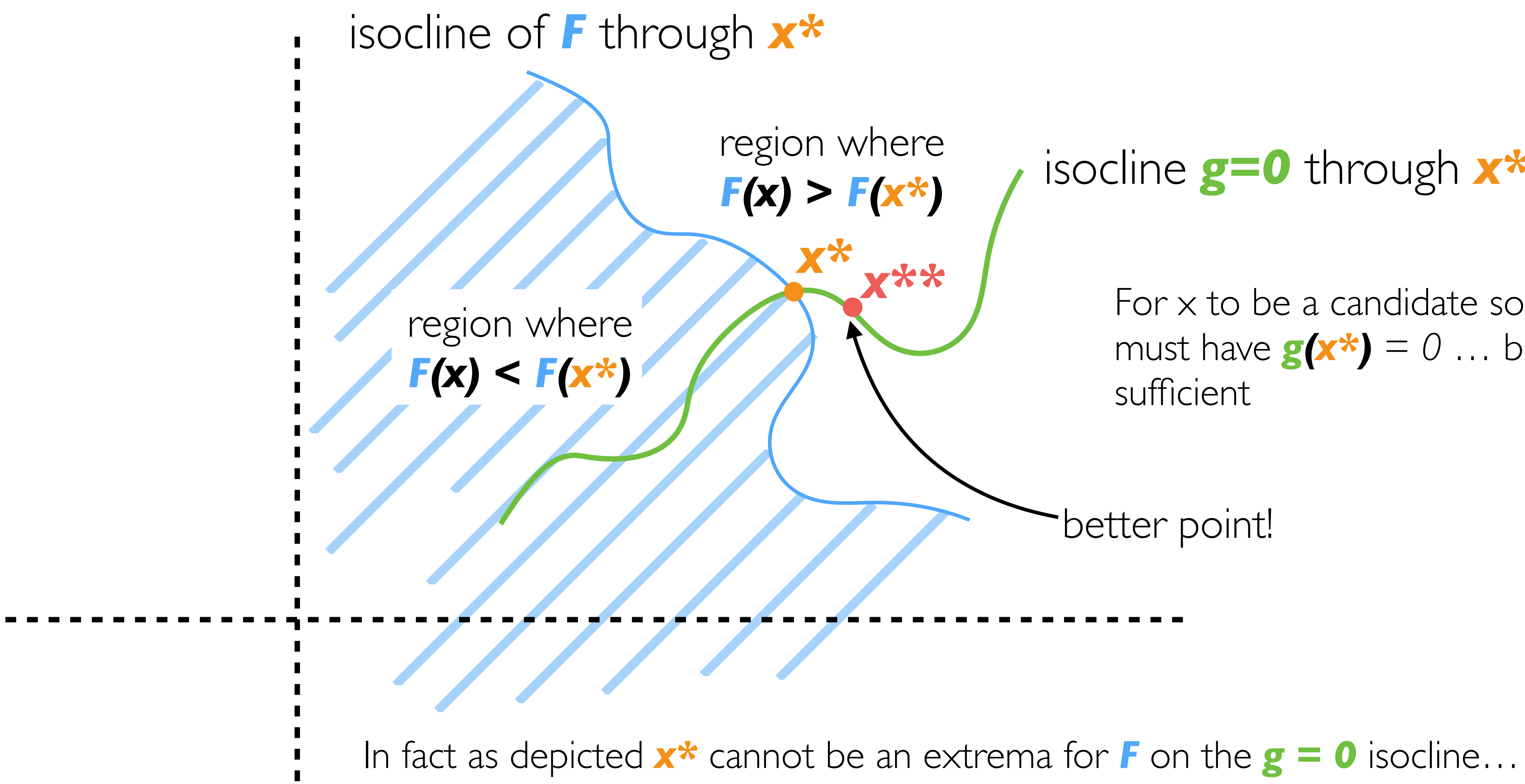

For x to be a candidate solution must have ***g*(*x**)** = *0* … but not sufficient

In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g* = *0*** isocline… why?

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

So: <u>true optima ***x**** must have isocline of ***F*** through ***x**** be parallel to ***g = 0*** isocline</u>



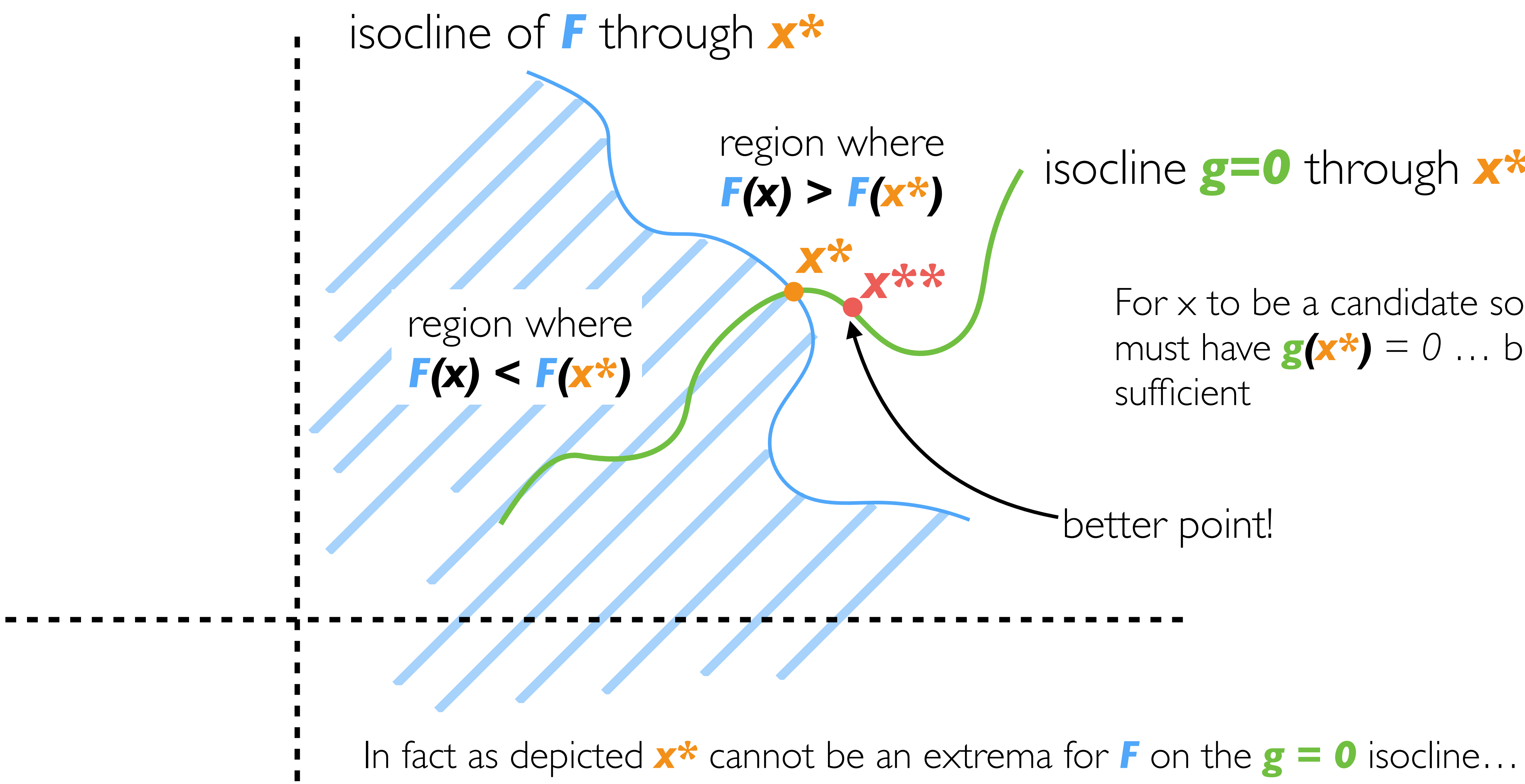

For x to be a candidate solution must have ***g***(***x****) = *0* … but not sufficient

In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g = 0*** isocline… why?

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g* = *0***?

So: <u>true optima ***x**** must have isocline of ***F*** through ***x**** be parallel to ***g* = *0*** isocline</u>




For x to be a candidate solution must have $g(x^*) = 0$ … but not sufficient

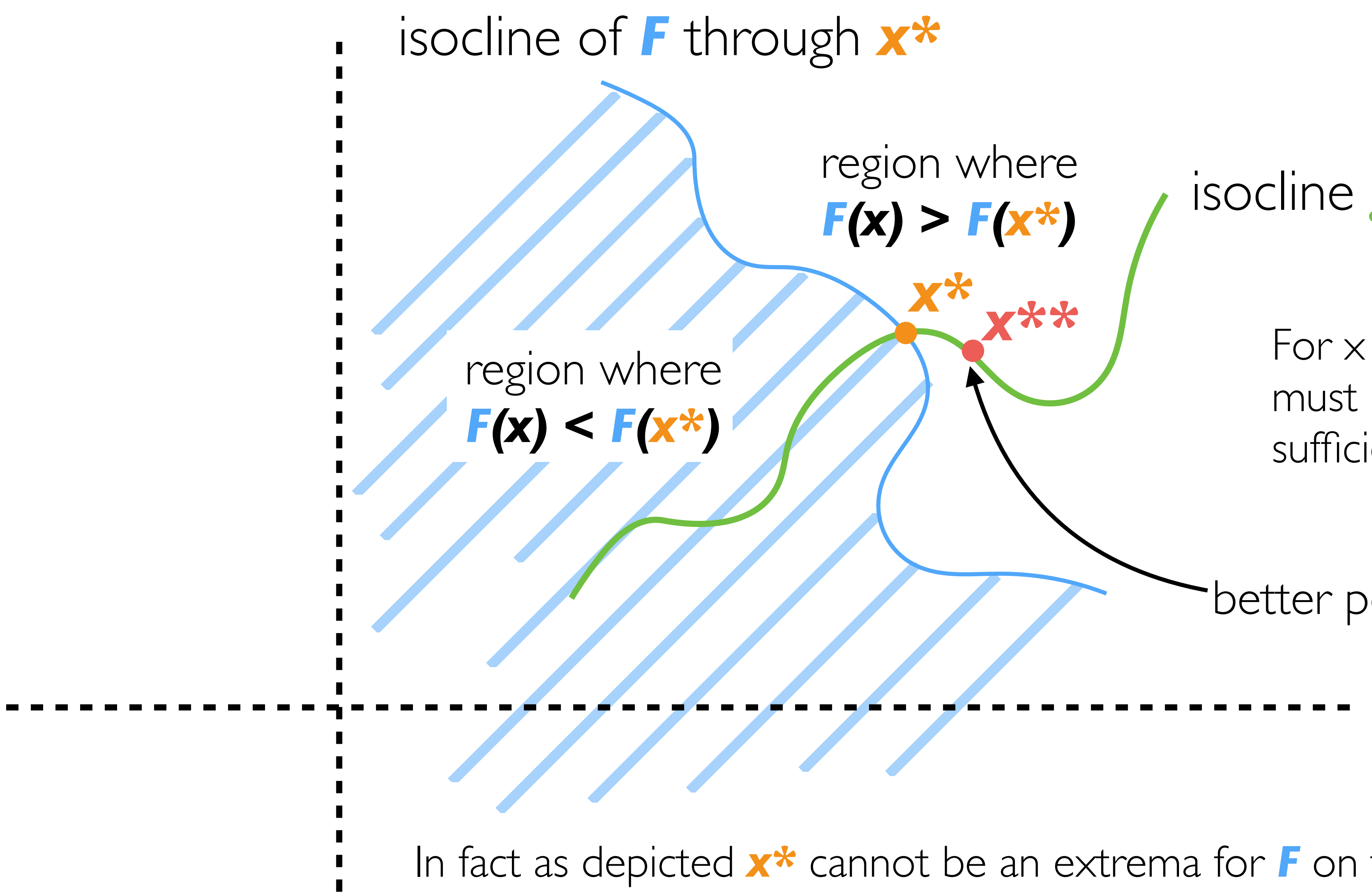
In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g* = *0*** isocline… why?

But if isoclines are parallel then so are gradient vectors, since gradients are perpendicular to the isoclines

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g* = *0***?

So: <u>true optima ***x**** must have isocline of ***F*** through ***x**** be parallel to ***g* = *0*** isocline</u>




For x to be a candidate solution must have ***g*(*x**)** = *0* … but not sufficient

In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g* = *0*** isocline… why?

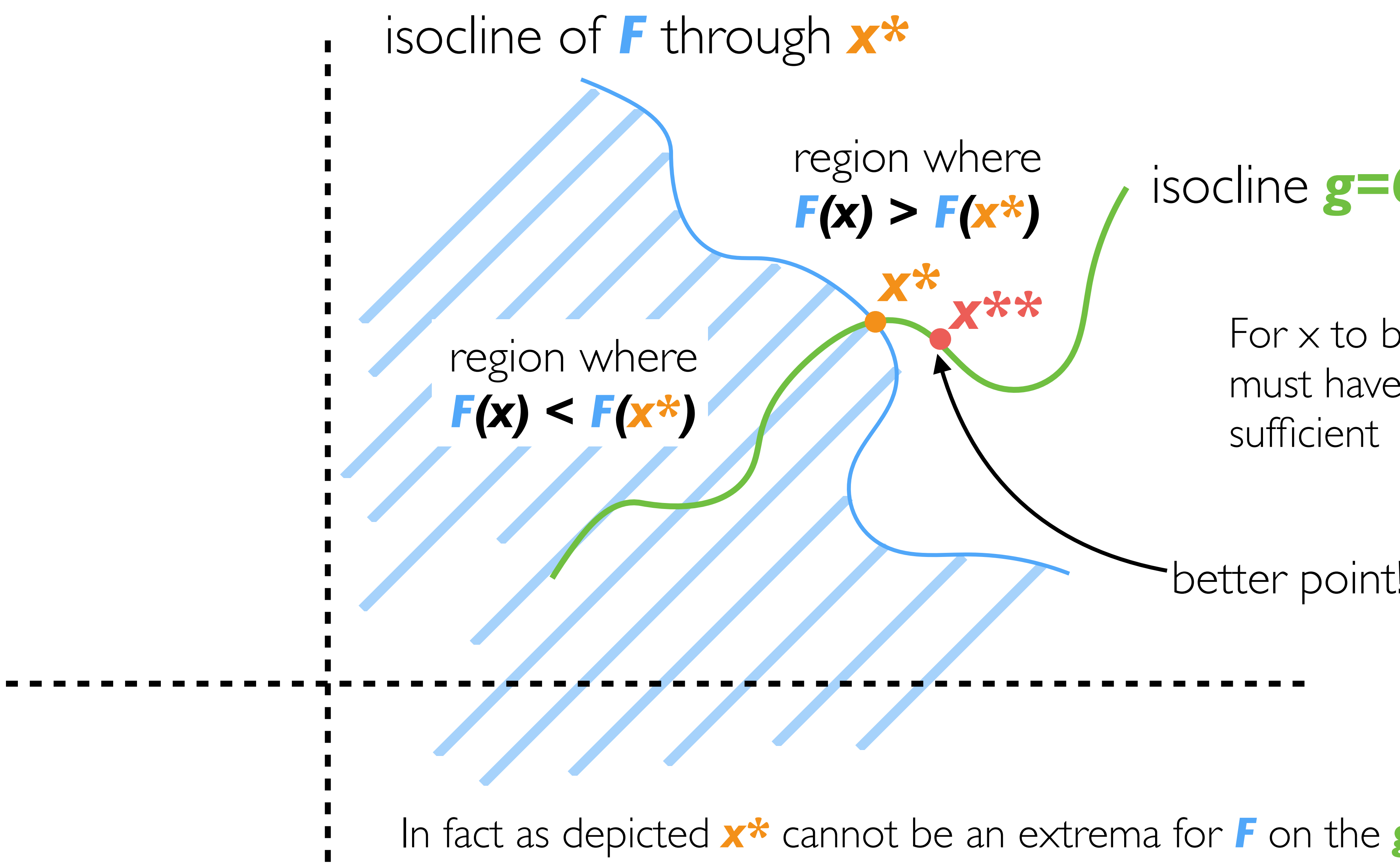
But if isoclines are parallel then so are gradient vectors, since gradients are perpendicular to the isoclines

$$\nabla F(x^*) \quad || \quad \nabla g(x^*)$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

So: <u>true optima ***x**** must have isocline of ***F*** through ***x**** be parallel to ***g = 0*** isocline</u>




For x to be a candidate solution must have **g(x*)** = *0* … but not sufficient

In fact as depicted ***x**** cannot be an extrema for ***F*** on the ***g = 0*** isocline… why?

But if isoclines are parallel then so are gradient vectors, since gradients are perpendicular to the isoclines

$$\nabla F(x^*) \quad || \quad \nabla g(x^*) \qquad \Rightarrow \qquad \nabla F(x^*) = \lambda \nabla g(x^*)$$

scalar

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x****** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - r^2)$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g* = *0***?

Find **λ**, ***x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - r^2)$$

$$\nabla \mathcal{L}(x_1, x_2, \lambda) = (1 - 2\lambda x_1, 1 - 2\lambda x_2, r^2 - x_1^2 - x_2^2)$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - r^2)$$

$$\nabla \mathcal{L}(x_1, x_2, \lambda) = (1 - 2\lambda x_1, 1 - 2\lambda x_2, r^2 - x_1^2 - x_2^2)$$

$$= 0 \quad \Rightarrow \quad x_1 = \frac{1}{2\lambda}, x_2 = \frac{1}{2\lambda}, \lambda = \pm\frac{1}{\sqrt{2}r}$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ**, ***x****** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$

$$\mathcal{L}(x_1, x_2, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - r^2)$$

$$\nabla \mathcal{L}(x_1, x_2, \lambda) = (1 - 2\lambda x_1, 1 - 2\lambda x_2, r^2 - x_1^2 - x_2^2) = 0$$

$$\Rightarrow \quad \vec{x}^* = \left(\frac{r}{\sqrt{2}}, \frac{r}{\sqrt{2}}\right) \text{ or } \left(-\frac{r}{\sqrt{2}}, -\frac{r}{\sqrt{2}}\right)$$

# Multi-variable Functions

What if our goal is to maximize ***F*** subject to the constraint that ***g = 0***?

Find **λ, *x**** such that $\nabla F(x^*) = \lambda \nabla g(x^*)$

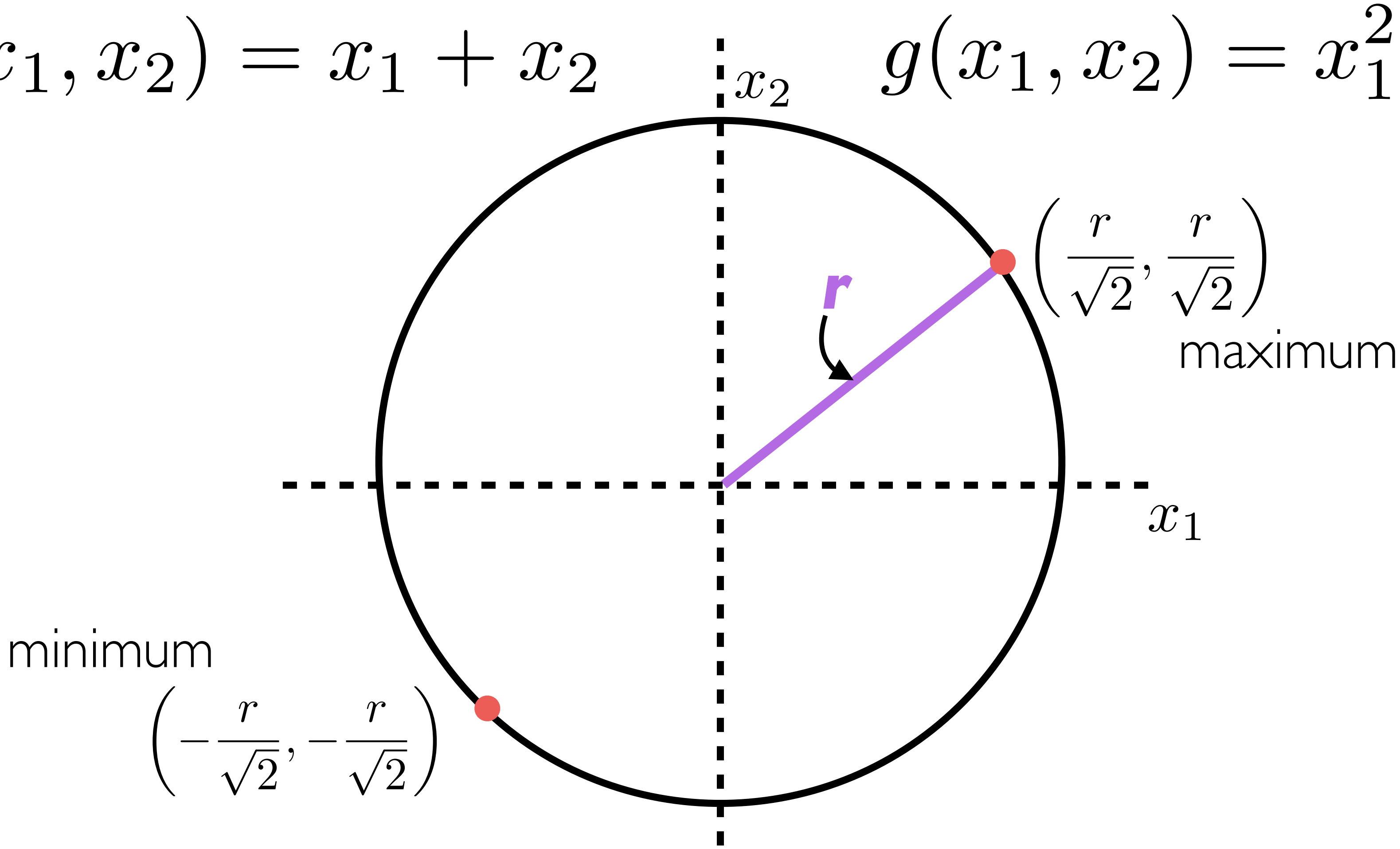
Essentially equivalent to maximizing: $\mathcal{L}(\vec{x}, \lambda) = F(\vec{x}) - \lambda g(\vec{x})$

E.g. suppose we want to maximize

$$F(x_1, x_2) = x_1 + x_2$$

Subject to:

$$g(x_1, x_2) = x_1^2 + x_2^2 - r^2 = 0$$



# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

What about multi-variable case?

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

What about multi-variable case?

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

*def'n of partial derivative*

$$\frac{\partial H(\vec{x})}{\partial x_1} = \lim_{\Delta x_1 \to 0} \frac{H(x_1 + \Delta x_1, x_2, \ldots) - H(x_1, x_2, \ldots)}{\Delta x_1}$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

What about multi-variable case?

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

*def'n of partial derivative*

$$\frac{\partial H(\vec{x})}{\partial x_1} = \lim_{\Delta x_1 \to 0} \frac{H(x_1 + \Delta x_1, x_2, \ldots) - H(x_1, x_2, \ldots)}{\Delta x_1}$$

*def'n of **H***

$$= \lim_{\Delta x_1 \to 0} \frac{G(F_1(x_1 + \Delta x_1, x_2, \ldots), \ldots, F_m(x_1 + \Delta x_1, x_2, \ldots)) - H(\vec{x})}{\Delta x_1}$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

What about multi-variable case?

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

*def'n of partial derivative*

$$\frac{\partial H(\vec{x})}{\partial x_1} = \lim_{\Delta x_1 \to 0} \frac{H(x_1 + \Delta x_1, x_2, \ldots) - H(x_1, x_2, \ldots)}{\Delta x_1}$$

*def'n of* ***H***

$$= \lim_{\Delta x_1 \to 0} \frac{G(F_1(x_1 + \Delta x_1, x_2, \ldots), \ldots, F_m(x_1 + \Delta x_1, x_2, \ldots)) - H(\vec{x})}{\Delta x_1}$$

*applying single-variable chain rule* ***m*** *times*

$$= \sum_i \frac{\partial G}{\partial y_i}(\vec{F}(\vec{x}))\frac{\partial F_i}{\partial x_1}(\vec{x})$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad h'(x) = g'(f(x)) \cdot f'(x)$$

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

So in general:

$$\frac{\partial H}{\partial x_j} = \sum_i \frac{\partial G}{\partial y_i}(\vec{F}(\vec{x}))\frac{\partial F_i}{\partial x_j}(\vec{x})$$

# Multi-variable Functions

Single-variable chain rule says that for:

$$h(x) = g(f(x)) \qquad\qquad h'(x) = g'(f(x)) \cdot f'(x)$$

$$H(\vec{x}) = G(F_1(\vec{x}), \ldots, F_m(\vec{x}))$$

So in general:

$$\frac{\partial H}{\partial x_j} = \sum_i \frac{\partial G}{\partial y_i}(\vec{F}(\vec{x}))\frac{\partial F_i}{\partial x_j}(\vec{x})$$

Or:

$$\nabla H = \nabla \vec{F}(\vec{x}) \cdot \nabla G(\vec{F}(\vec{x}))$$

***n***-vector-valued fn ***nxm-***matrix-valued fn ***m-***vector-valued fn

Where $\nabla \vec{F}$ is the matrix with ***ij***-th elt:

$$(\nabla \vec{F})_{ji} = \frac{\partial F_i}{\partial x_j}$$

# Multi-variable Functions

Taylor series for function ***f(x1, ... xn)*** of ***n*** inputs:

pointwise mult.

$$f(\vec{x}_0, \Delta\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) * \Delta\vec{x} + \frac{1}{2}(H[F] \cdot \vec{x}_0) * \Delta x^2 + \ldots$$

vector delta

matrix-on-vector multiplication

Hessian matrix

$$H[f] = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \frac{\partial^2 f}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

shape = (***n***, ***n***)