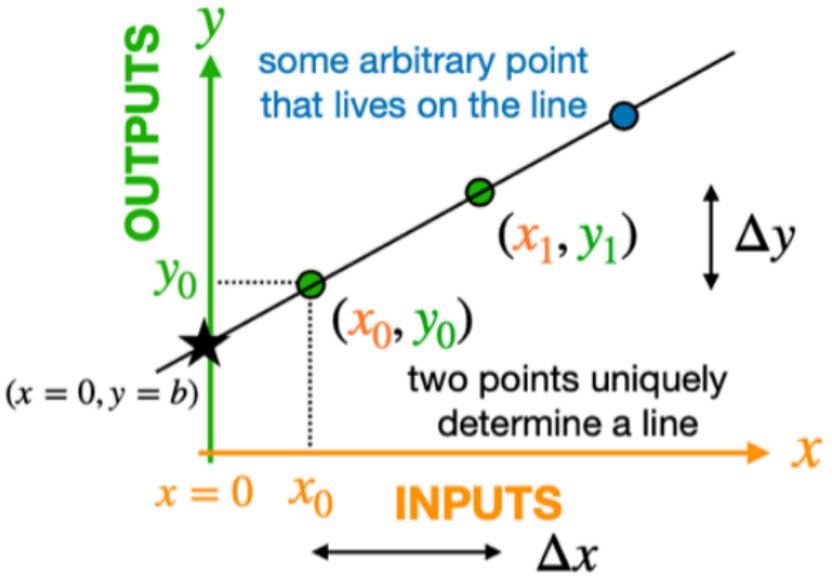
Linear functions



Standard form: $Ax + By = C$

Two Parameter model:

slope $m = \frac{rise}{run} = \frac{\Delta y}{\Delta x} = \frac{y_1 - y_0}{x_1 - x_0}$

y-Intercept: $x = 0 \rightarrow f(0) = b$

Common forms:

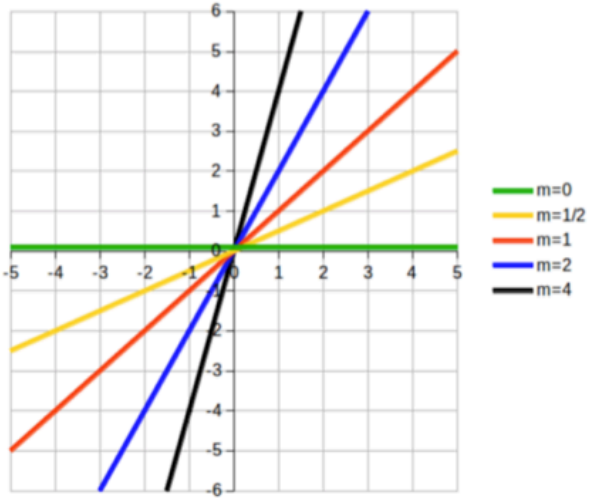
slope-intercept form:

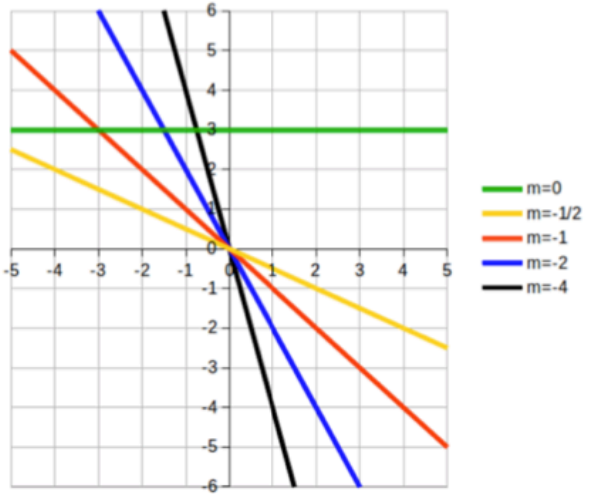
$$y = f(x) = mx + b \qquad \mathbf{P} = (m, b)$$

Point-slope form:

$$(y - y_0) = m(x - x_0) = \left(\frac{y_1 - y_0}{x_1 - x_0}\right)(x - x_0)$$

Meaning of slope:





Function parameterization

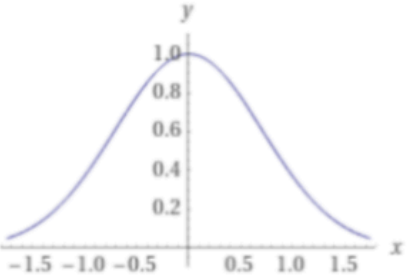
- This is VERY important in parametric statistical learning.
- **FOR PARAMETRIC MODELS, THE FUNCTION PARAMETERS ARE WHAT WE LEARN!!**
- Think of parameters as “knobs” we can turn to bend and shift the function “fit” the data

PARAMETRIC MODEL = Functional form + Parameterization

(underlying shape) (stretching and shifting)

$y = f(x) \rightarrow$ **"parent function"**

$$f(x) = e^{-x^2}$$



Parameterized form-1:

$$y = f(x \mid \mathbf{p}) = Af\left(\frac{x - x_o}{w}\right) + S$$

$$\mathbf{p} = (A, x_o, w, S)$$

A = **verticle stretch (amplitude)**

x_o = **horizontal shift (recentering parameter)**

w = **horizontal stretch (width parameter)**

S = **verticle shift (shift parameter)**

Parameterized form-2: $y = f(x \mid \mathbf{p}) = Af(w_1 x + b) + S$

(weights and bias) $w_1 = \frac{1}{w}$ **and** $b = -\frac{x_o}{w}$

Function transformations

Transformation Rules for Functions		
Function Notation	Type of Transformation	Change to Coordinate Point
f(x) + d	Vertical translation up d units	(x, y) → (x, y + d)
f(x) − d	Vertical translation down d units	(x, y) → (x, y − d)
f(x + c)	Horizontal translation left c units	(x, y) → (x − c, y)
f(x − c)	Horizontal translation right c units	(x, y) → (x + c, y)
−f(x)	Reflection over x-axis	(x, y) → (x, −y)
f(−x)	Reflection over y-axis	(x, y) → (−x, y)
af(x)	Vertical stretch for |a| > 1	(x, y) → (x, ay)
	Vertical compression for 0 < |a| < 1	
f(bx)	Horizontal compression for |b| > 1	$(x, y) \to \left(\frac{x}{b}, y\right)$
	Horizontal stretch for 0 < |b| < 1	

Rigid

Non-Rigid

https://www.onlinemathlearning.com/parent-functions.html

Parameterized form-1:

$$y = f(x \mid \mathbf{p}) = Af\left(\frac{x - x_o}{w}\right) + S$$

$$\mathbf{p} = (A, x_o, w, S)$$

A = verticle stretch (amplitude)

x_o = horizontal shift (recentering parameter)

w = horizontal stretch (width parameter)

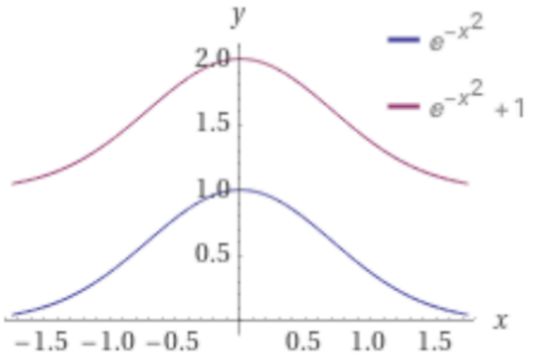
S = verticle shift (shift parameter)

Parameterized form-2: $y = f(x \mid \mathbf{p}) = Af(w_1 x + b) + S$

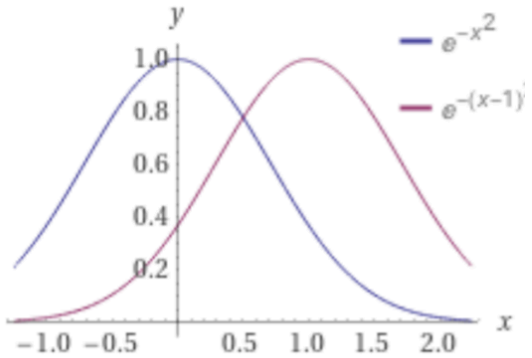
(weights and bias) $w_1 = \frac{1}{w}$ **and** $b = -\frac{x_o}{w}$

- Demonstration of a function getting bent and shifted via parametrization

addition outside (rigid verticle shift): $f(x)+a$

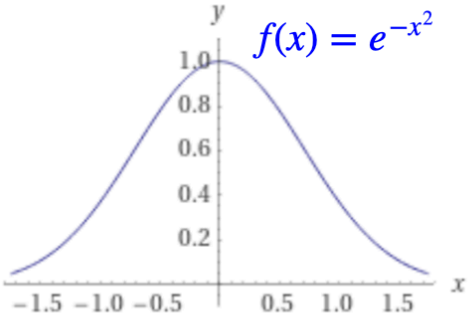


addition inside (rigid horizontal shift): $f(x-a)$

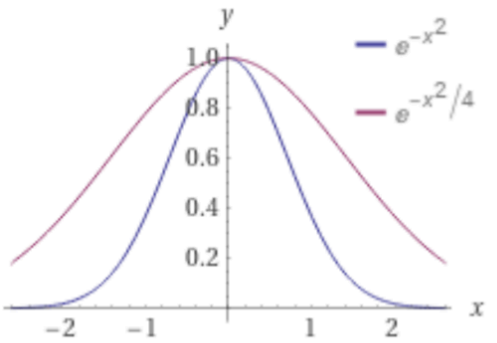


Parent function:

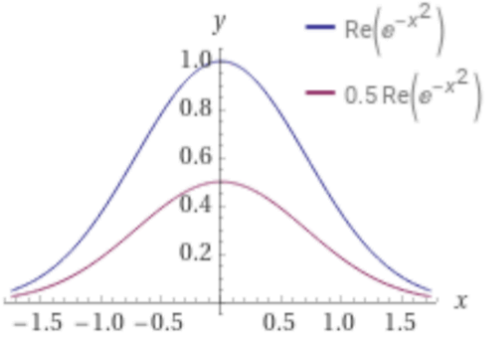
$f(x)=e^{-x^2}$



multiplication inside (non-rigid horizontal scaling): $f(ax)$



multiplication outside (non-rigid verticle scaling): $af(x)$



Additional Examples

- The following are common single variable parent functions

(underlying shape)

$y = f(x) \rightarrow$ **"parent function"**

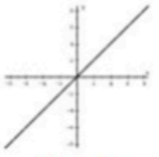
(stretching and shifting) $\mathbf{p} = (A, x_o, w, S)$

$$y = f(x \mid \mathbf{p}) = Af\left(\frac{x - x_o}{w}\right) + S$$

Linear

$$y = f(x \mid \mathbf{p}) = mx + b$$

Linear

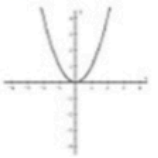


$\mathbf{p} = (m, b)$

Quadratic

$$y = f(x \mid \mathbf{p}) = Ax^2 + Bx + C$$

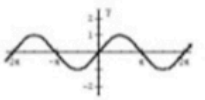
Quadratic regressions



$\mathbf{p} = (A, B, C)$

Sinusoidal

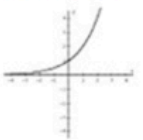
$$y = f(x \mid \mathbf{p}) = A_o \sin(\omega t + \phi) + c$$



$\mathbf{p} = (A_o, \omega, \phi, c)$

Exponential

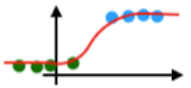
$$y = f(x \mid \mathbf{p}) = ae^{\frac{(x - x_o)}{w}} + d$$



$\mathbf{p} = (a, w, x_o, d)$

$$y = f(x \mid \mathbf{p}) = \frac{A}{1 + e^{-\frac{(x - x_o)}{w}}} + s$$

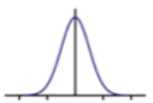
Logistic



$\mathbf{p} = (A, w, x_o, S)$

$$y = f(x \mid \mathbf{p}) = A \cdot \exp^{-\left(\frac{x - x_o}{w}\right)^2} + S$$

Gaussian



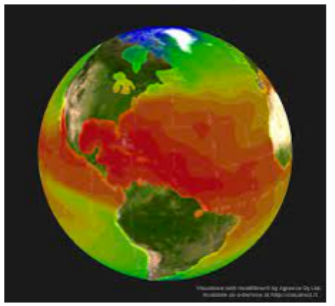
$\mathbf{p} = (A, x_o, w, S)$

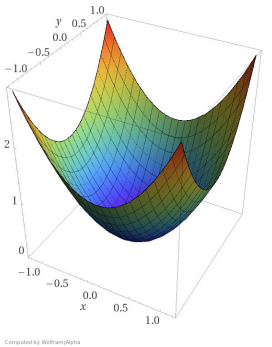
Multi-variable function types

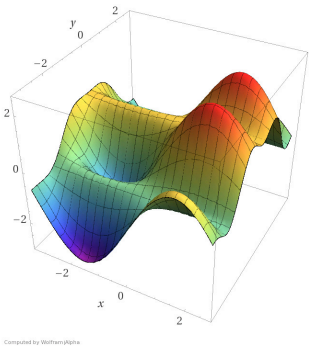
- **Scalar fields:** $y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_N) \qquad \mathbb{R}^N \rightarrow \mathbb{R}^1.$
  - **IMPORTANT:** In regression, the objective function is called the loss function L (or loss surface). For parametric modeling, the loss surface is a scalar field over the model parameters. (e.g. $L(\mathbf{p})$ or $L(\mathbf{w})$ depending on the parameter notation).

vector input: $\mathbf{x}$
scalar output: y

$T = T(\theta, \phi)$

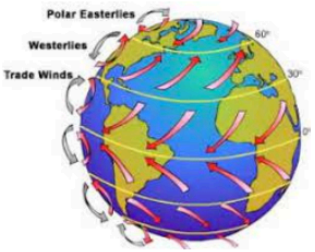


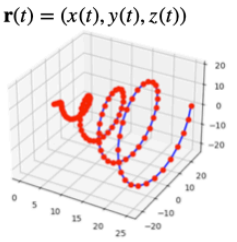




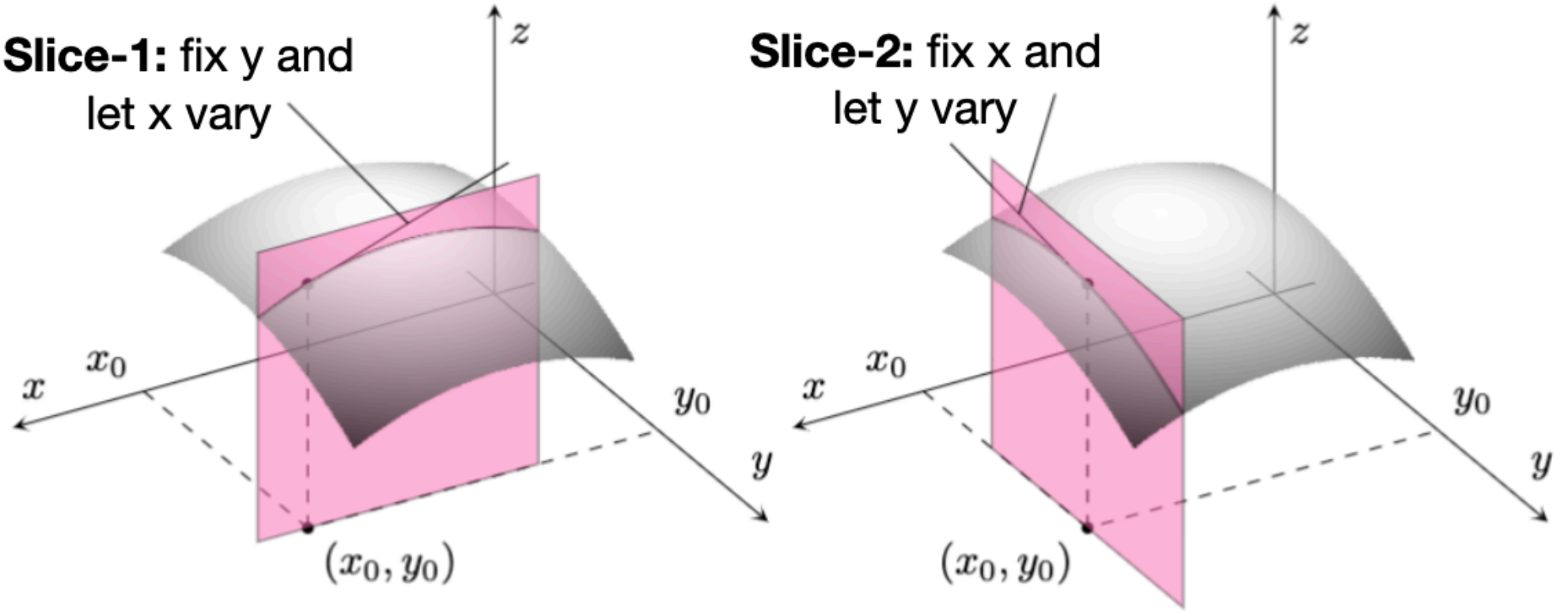
- **Vector fields:** $\mathbf{y} = f(\mathbf{x}) = \mathbf{f}(x_1, x_2, \ldots, x_N) = (y_1, y_2, \ldots, y_M) \qquad \mathbb{R}^N \rightarrow \mathbb{R}^M.$
  - **IMPORTANT:** The loss function's gradient $\nabla L(\mathbf{w})$ in regression is a vector field.

vector or scalar input: $\mathbf{x}$
vector output: $\mathbf{y}$



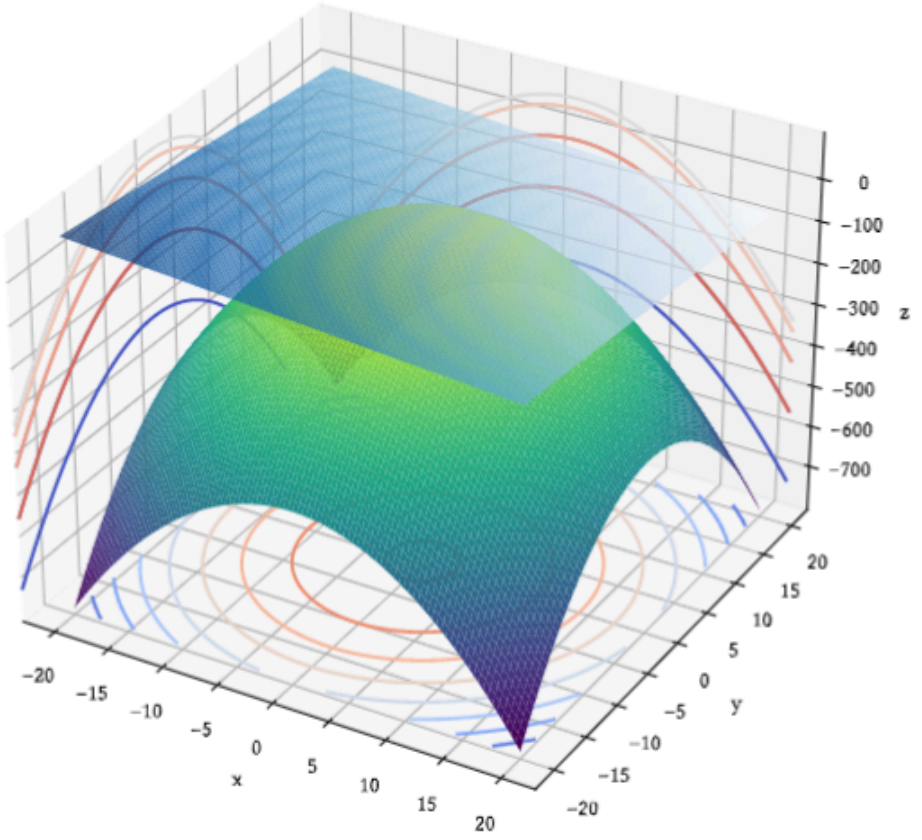


- Slicing is a very important concept in multi-variable calculus. It is when you “freeze” ALL variables, except one, and allow this variable to change independently
- This process converts a multi-variable scalar field $f(\mathbf{x}) = f(x_0, x_1 \ \ldots \ x_N)$, into a collection of single variable functions $f_1(x_1), f_2(x_2), \ldots f_N(x_N)$, in the neighborhood of some point in the multi-dimensional space $\mathbf{x}_0$. The various functions are single-variable because all other variables are “frozen” during their evaluation.



Contour plots-1

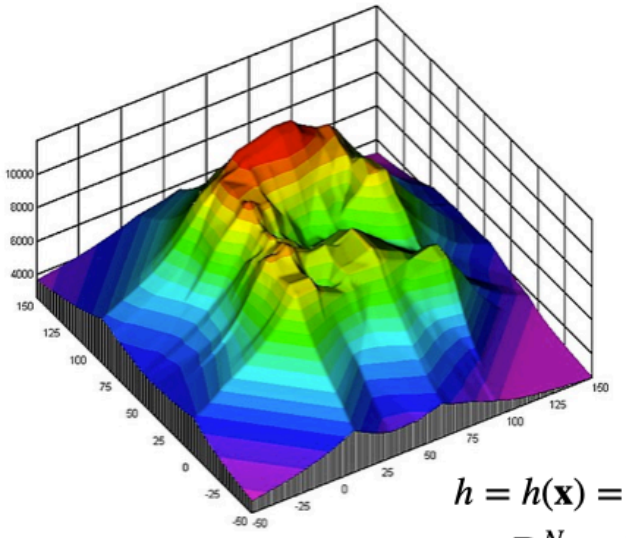
- Contour plots use slicing to create reduced dimensionality representations of higher dimensional data
- In this case, you freeze ONE variable, and visualize the effect of the variation in the others.



Contour plots-2

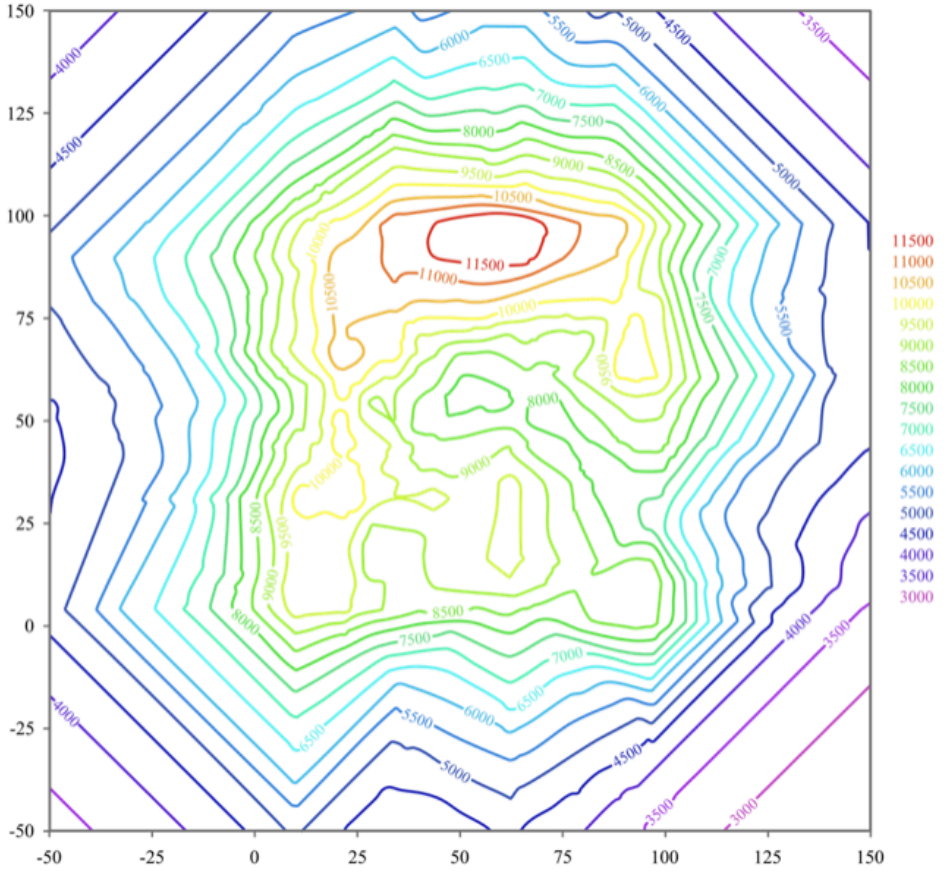
- Example of a scalar field with many local minima and maxima





$$h = h(\mathbf{x}) = h(x, y)$$

$$\mathbb{R}^N \rightarrow \mathbb{R}^1$$

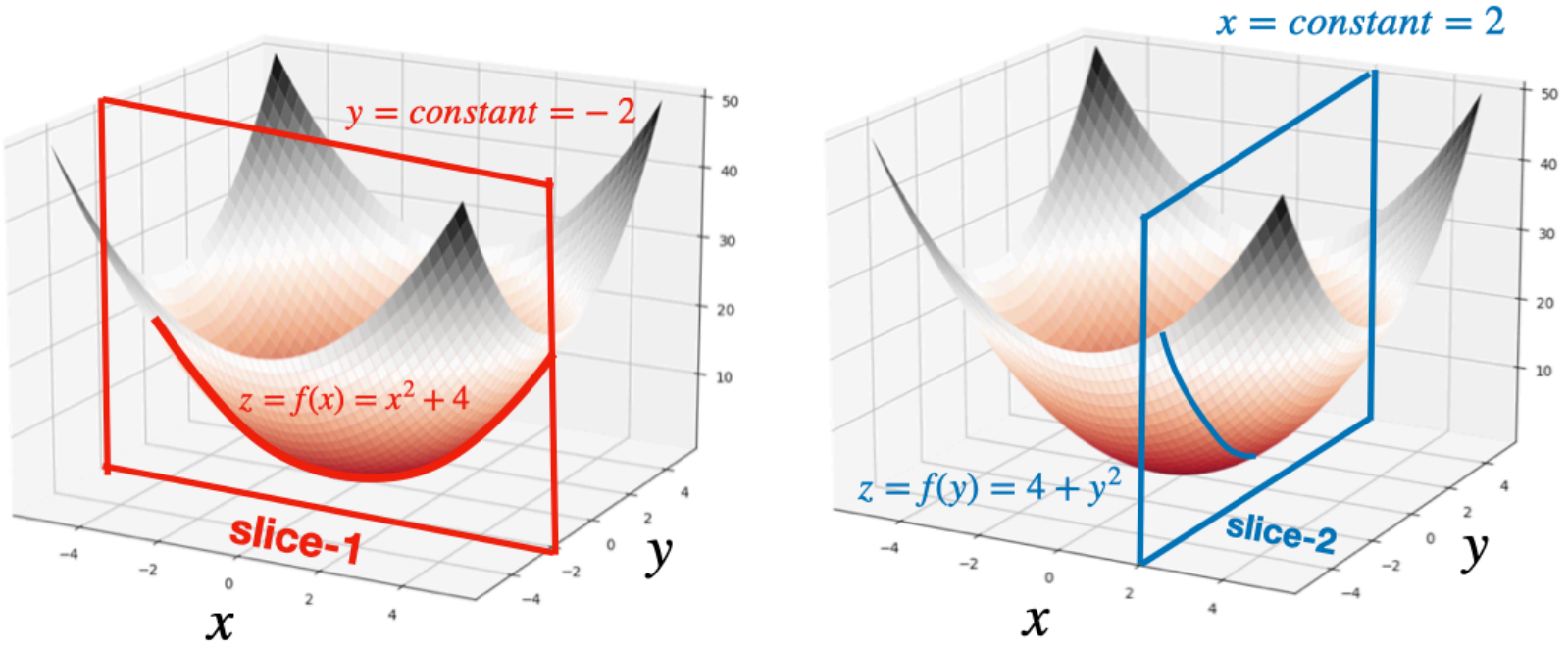


Contour plot

Fix h at constant intervals and let x and y vary

Example: Partial differentiation

$z = f(x, y) = x^2 + y^2$ **consider point:** $(x, y) = (2, -2)$



Partial derivative: x
hold y as constant and take derivative with respect to x

$$f_x(x, -2) = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, -2)$$

$$= \frac{\partial}{\partial x}\left(x^2 + 4\right) = 2x + 0 = 2x$$

Partial derivative: y
hold x as constant and take derivative with respect to y

$$f_y(2, y) = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(2, y)$$

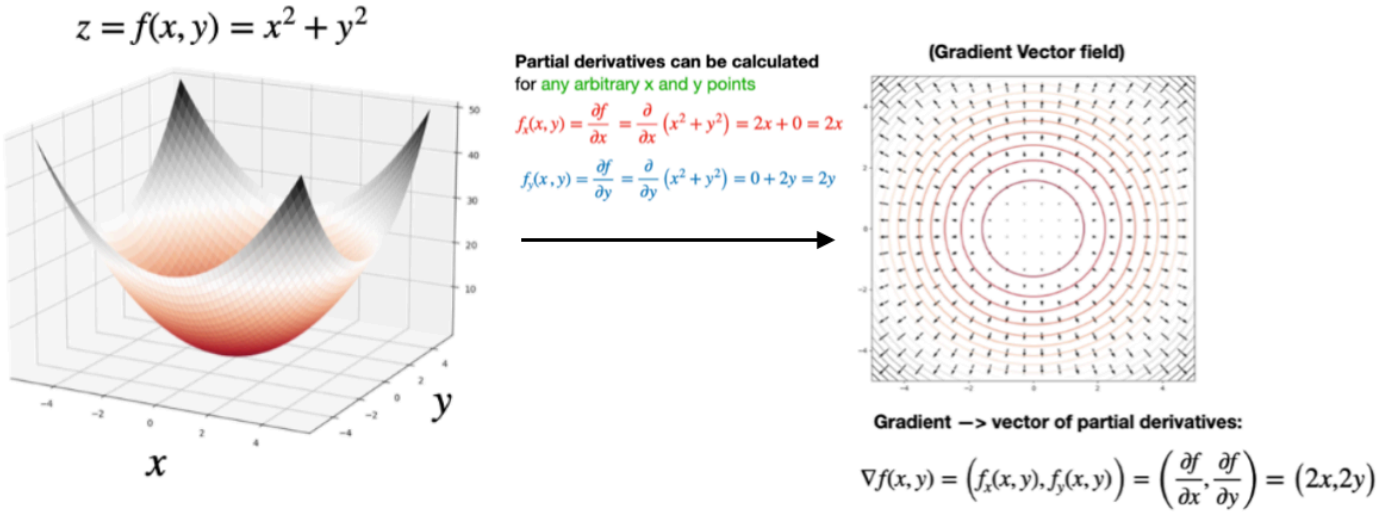
$$= \frac{\partial}{\partial y}\left(4 + y^2\right) = 0 + 2y = 2y$$

Partial derivatives at x=2 , y=-2

$$f_x(2, -2) = 2x\Big|_{x=2} = 4 \qquad f_y(2, -2) = 2y\Big|_{y=-2} = -4$$

The gradient

- The gradient is the vector of partial derivatives $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}, \frac{\partial f}{\partial w} \right)$



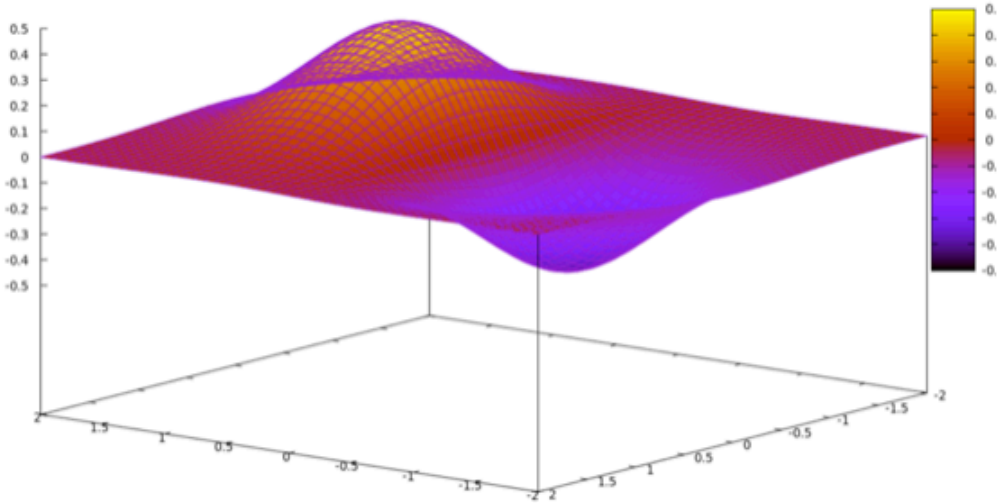
- **The gradient vector around a point $\mathbf{x}_0$ points in the direction of maximum increase of the function. This makes it incredibly useful for optimization**
- Gradient's magnitude is the rate of change in the direction of maximal increase
- The gradient can be thought of as an "operator" that acts on a function.

Gradient operator $\nabla = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix}$ → Acts on a scalar field $y = f(\mathbf{x}) = f(x_1, x_2, \ldots, x_N)$ → Gradient (Vector field) $\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$

Another example

- Minima or maxima correspond to special $\mathbf{x}$ where **ALL gradient components are zero**
- This is a necessary condition for find stationary points $\nabla f(\mathbf{x}) = \mathbf{0} = (0, 0, \ldots, 0)$

$$f(x, y) = xe^{-(x^2+y^2)}$$

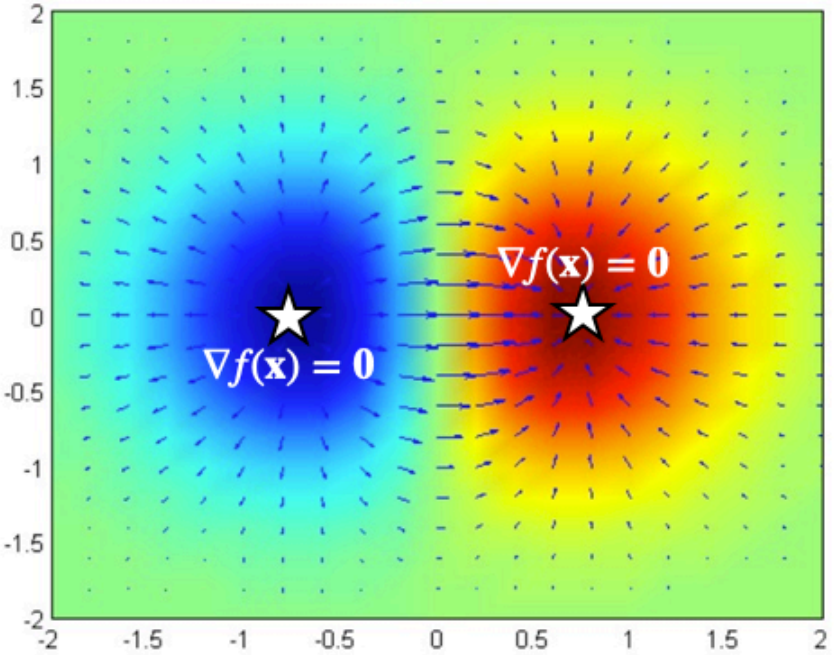


$$f_x(y, y) = \frac{\partial}{\partial x} f(x, y) = e^{-(x^2+y^2)}(1 - x\,(2x))$$

$$f_y(x, y) = \frac{\partial}{\partial y} f(x, y) = x\,(-2y)\,e^{-(x^2+y^2)}$$

product rule: $(u \cdot v)' = u' \cdot v + u \cdot v'$

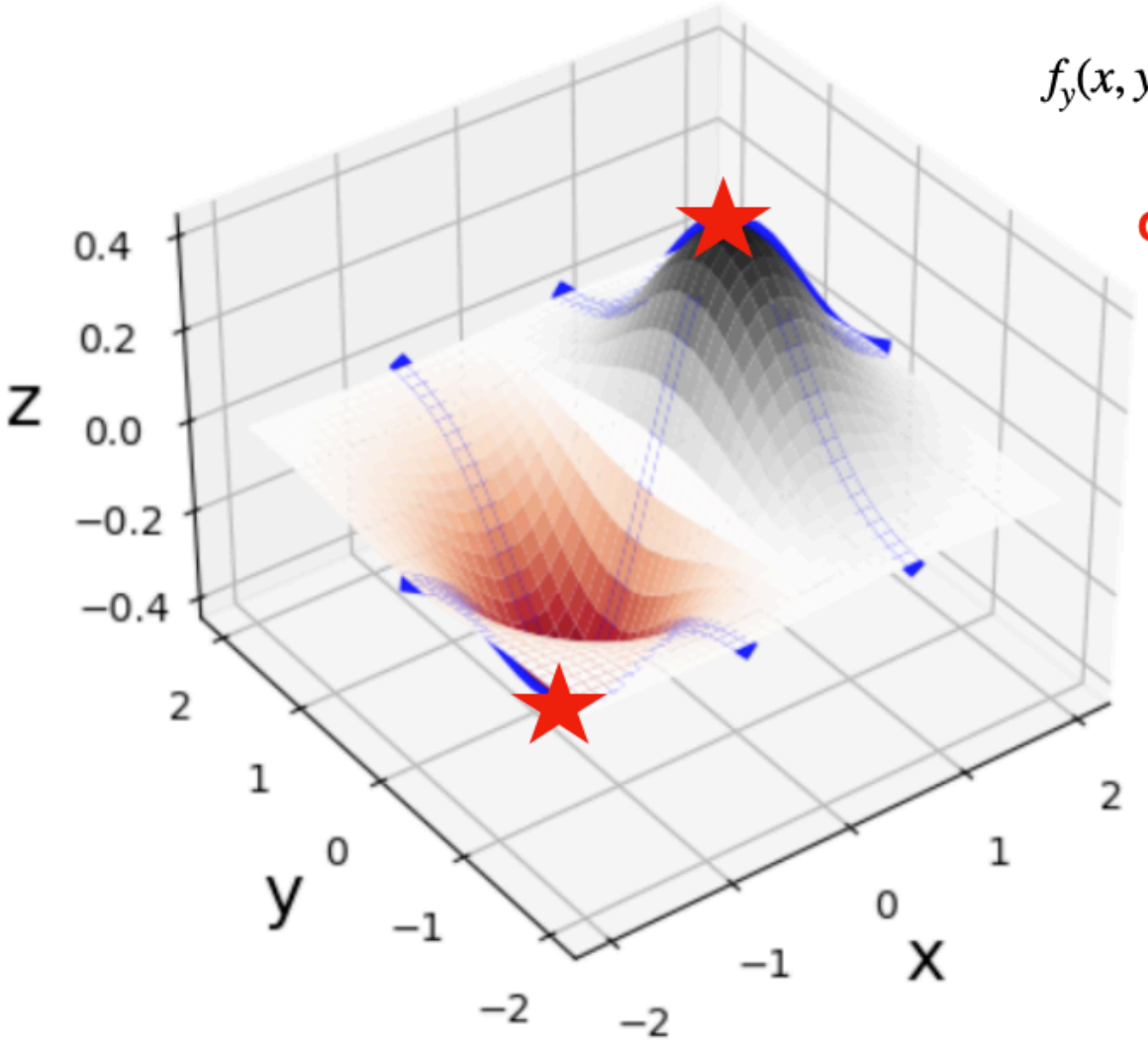
$$\nabla f(\mathbf{x}) = \left(f_x(x, y), f_y(x, y)\right) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$



(Gradient Vector field)

Multi-variable stationary points

$$f(x,y) = xe^{-(x^2+y^2)}$$



$$f_x(x,y) = \frac{\partial}{\partial x} f(x,y) = e^{-(x^2+y^2)}(1 - x\,(2x))$$

$$f_y(x,y) = \frac{\partial}{\partial y} f(x,y) = x\,(-2y)\,e^{-(x^2+y^2)}$$

Condition for minima and maxima

$$\boxed{\nabla f(\mathbf{x}) = \mathbf{0}}$$

Both terms of gradient need to be zero

Find the roots Of a system of EQ

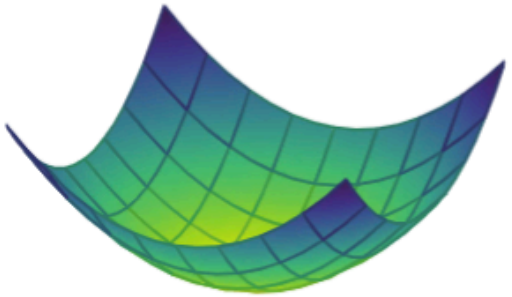
$$f_x(x,y) = 0$$

$$f_y(x,y) = 0$$

Multi-variable concavity

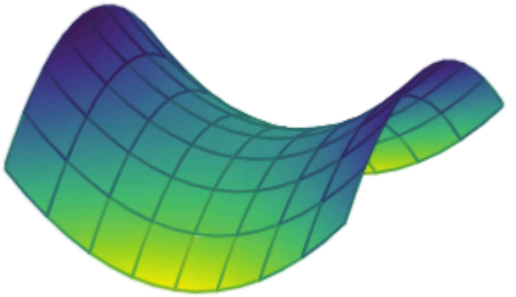
- Similar to the 1D case, the function's concavity control the type of stationary point

Strictly Convex



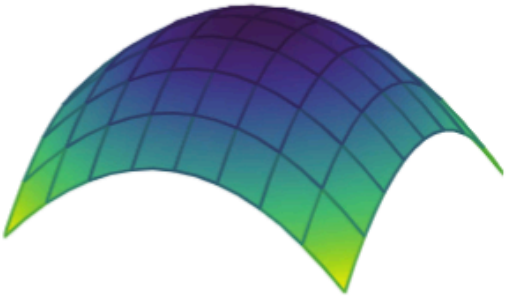
Global Minimum

Neither Concave nor Convex



Possibly a ***Saddle-Point*** or a Local Minimum / Maximum

Strictly Concave



Global Maximum