Neural Network Training:
Understanding Backpropagation requires an Understanding of Multivariate Functions, Derivates, and the General Chain Rule

Single Variate Derivative

$f(x)=6x^2$

$f'(x)=12x$ Lagrange notation

$df/dx = 12x$ Leibnitz notation

Single Variate Derivative

$f(x)=6x^2 = 3(2x^2) = 3\ g(x)$ nested function

Chain rule:

$g'(x)=4x$

$f'(x) = 3\ g'(x) = 3\ (\ 4\ x) = 12\ x$

Single Variate Derivative

$f(x)=6x^2 = 3(2x^2) = 3\ g(x)$ nested function

Chain rule:

$g'(x)=4x$

$f'(x) = 3\ g'(x) = 3 * (\ 4\ x) = 12\ x$

$df(g)/dx = df/dg * dg/dx$

Example of a 3D function and its partial derivatives:

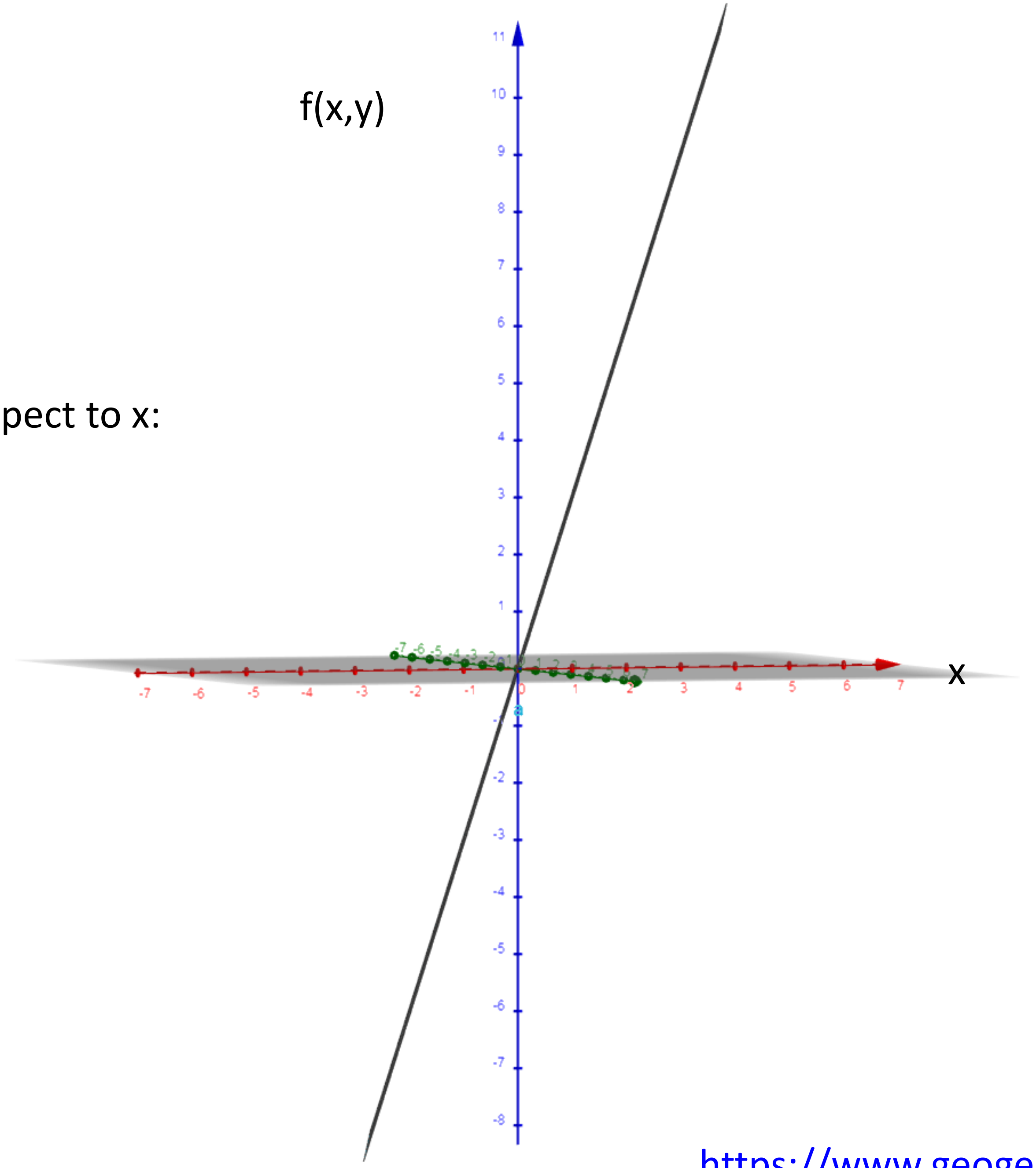
$f(x,y) = 3x+y$

$f(x,y)=3x+y$

Slope in x:

Derivative with respect to x:

$\partial f/\partial x \quad = 3$

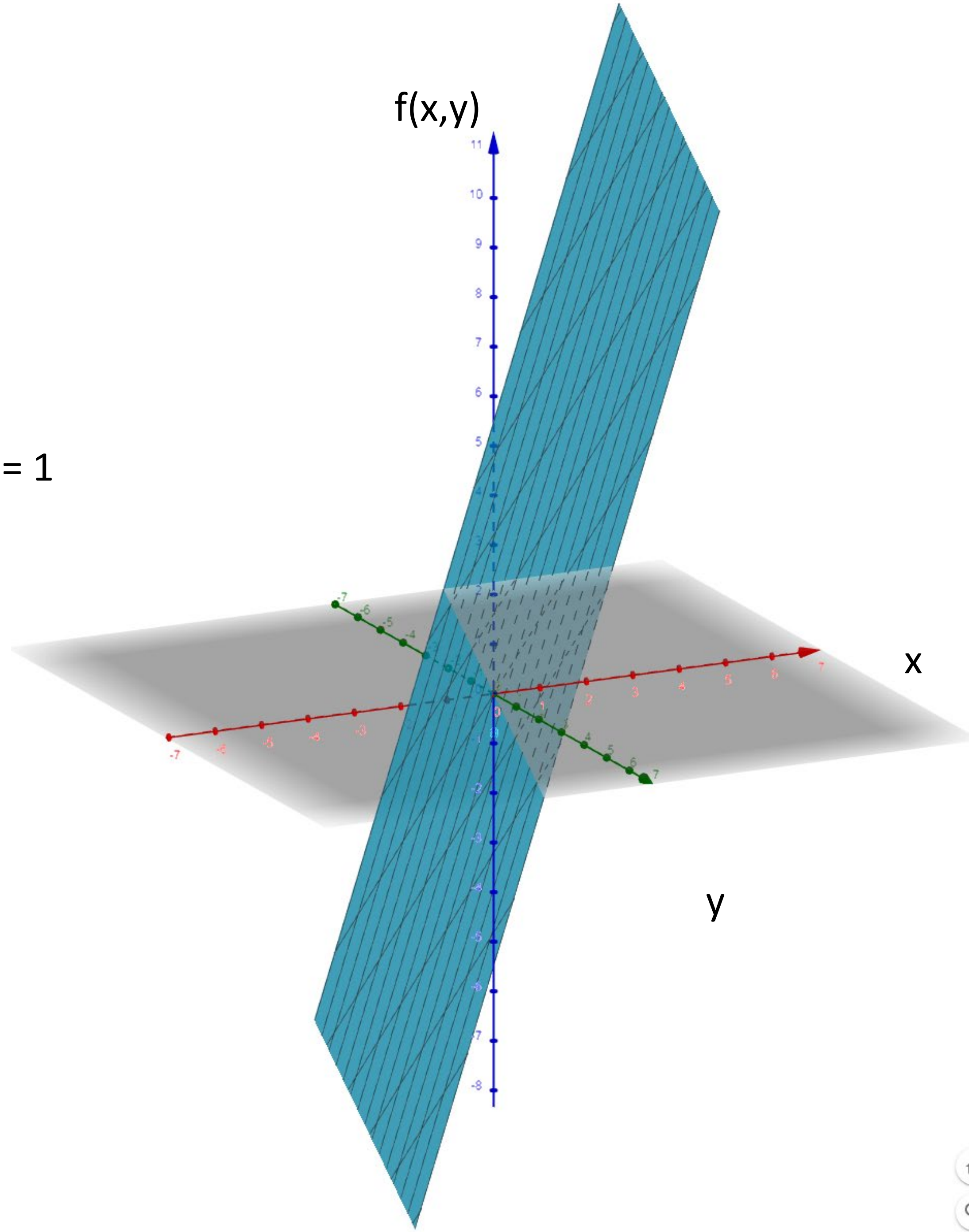


https://www.geogebra.org

f(x,y)=3x+y

Slope in y:

$\partial f/\partial y$ = 1



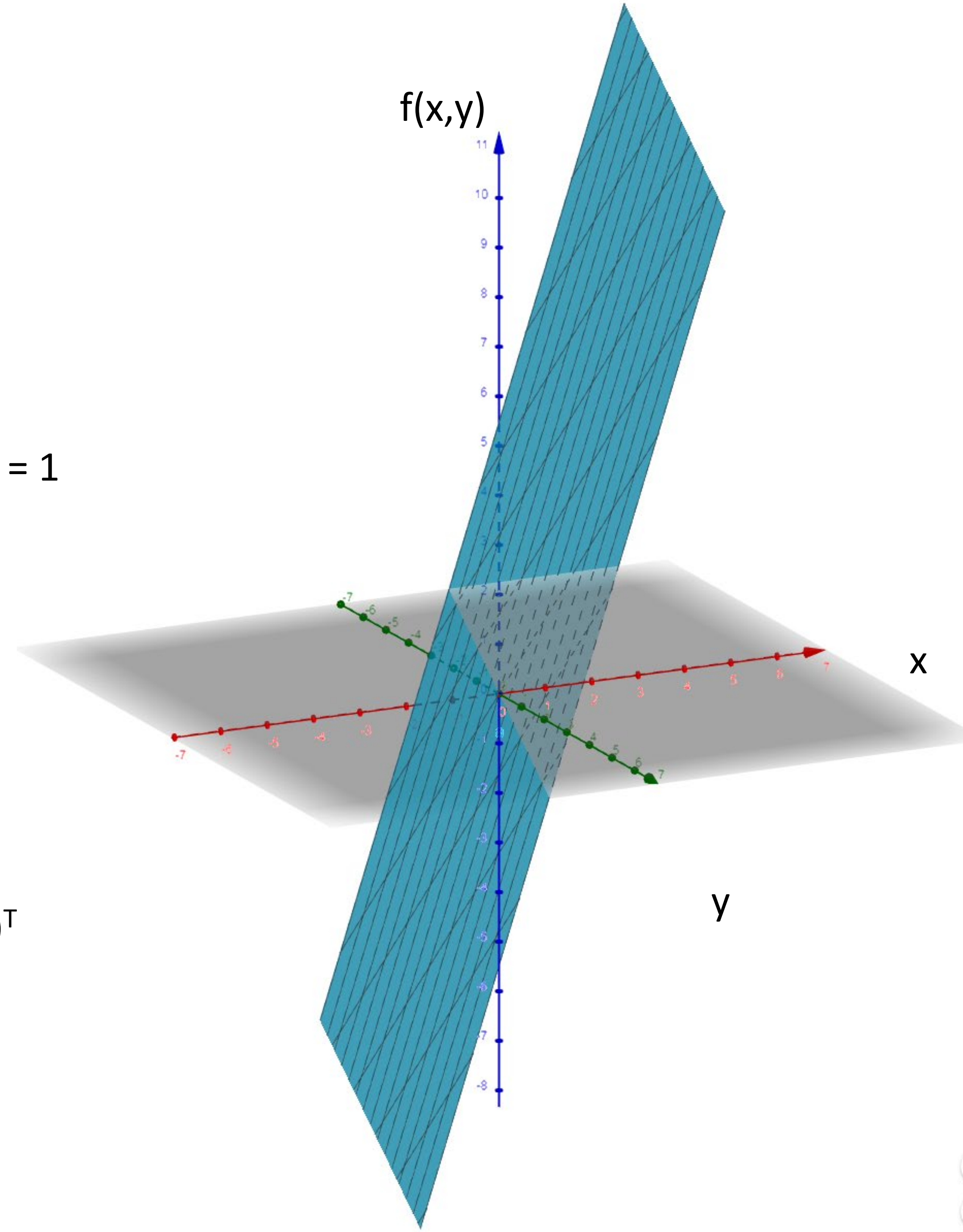
$f(x,y)=3x+y$

Slope in y:

$\partial f/\partial y = 1$

Gradient:

$\nabla f = (3, 1)^T$



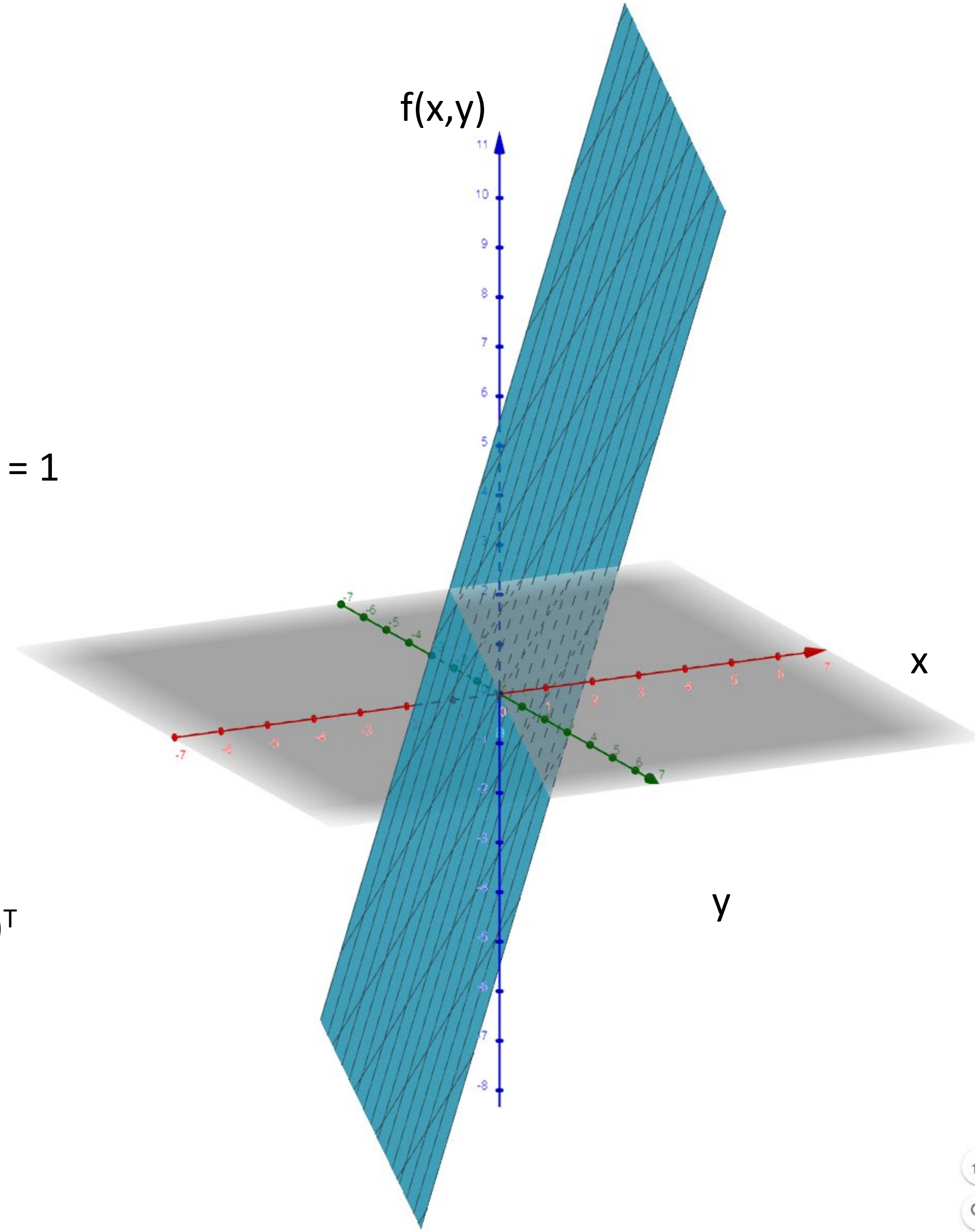
$f(x,y)=3x+y$

Slope in y:

$\partial f/\partial y \qquad = 1$

Gradient:

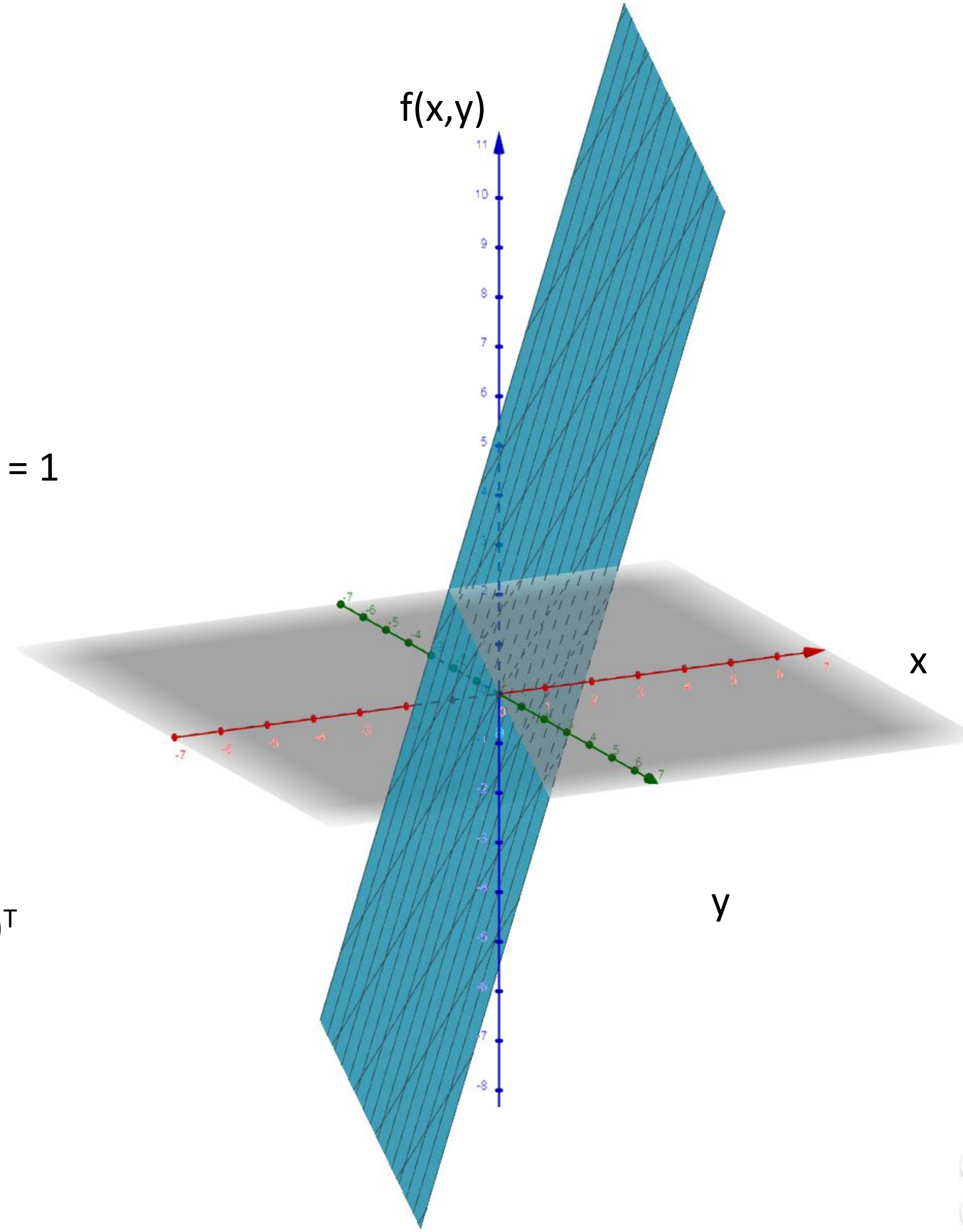
$\nabla f \quad = (3, 1)^T$



$f(x,y)=3x+y$

Slope in y:

$\partial f/\partial y \quad = 1$



Gradient:

$\nabla f \quad = (3, 1)^T$

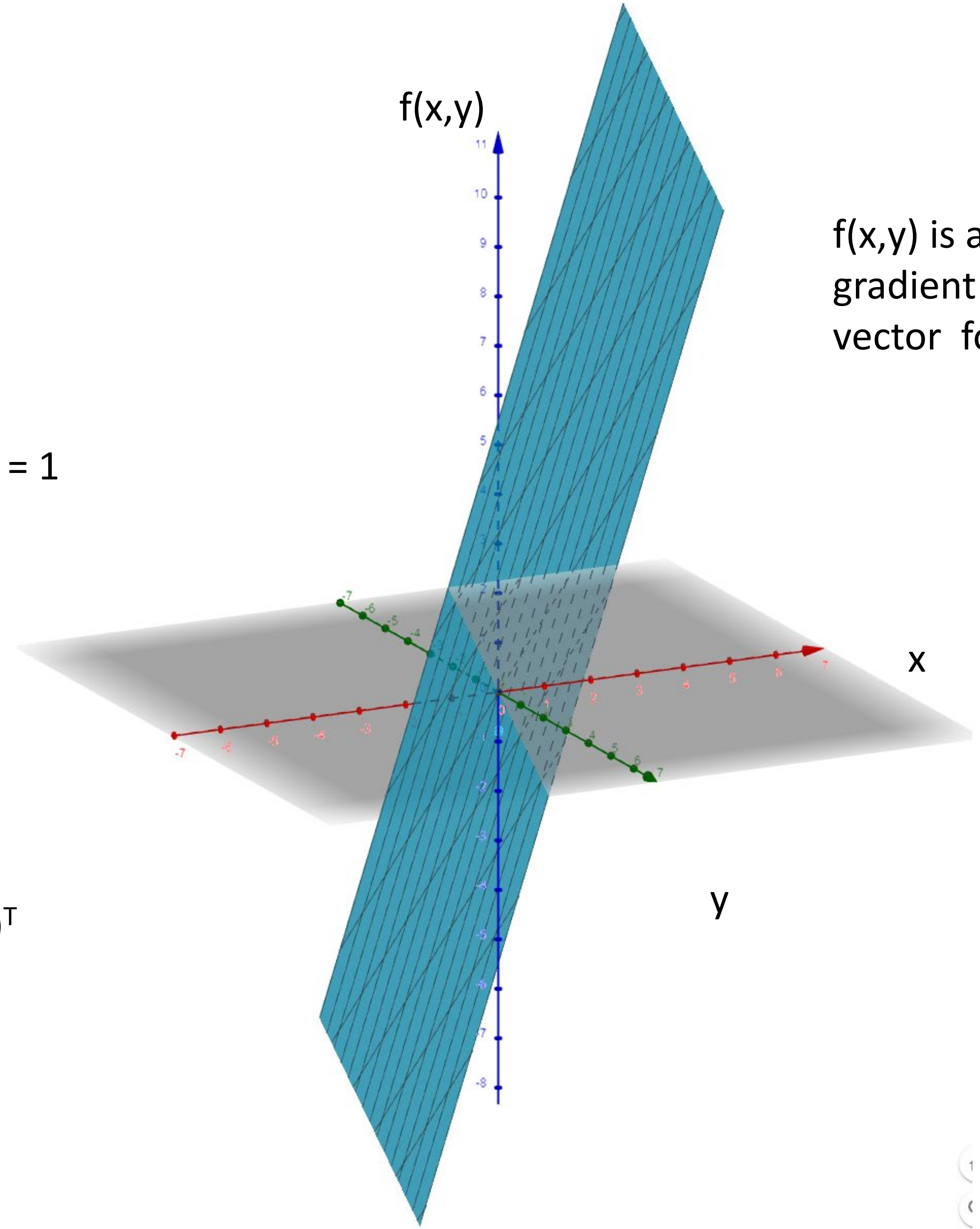
$f(x,y)=3x+y$

Slope in y:

$\partial f/\partial y = 1$

Gradient:

$\nabla f = (3, 1)^T$

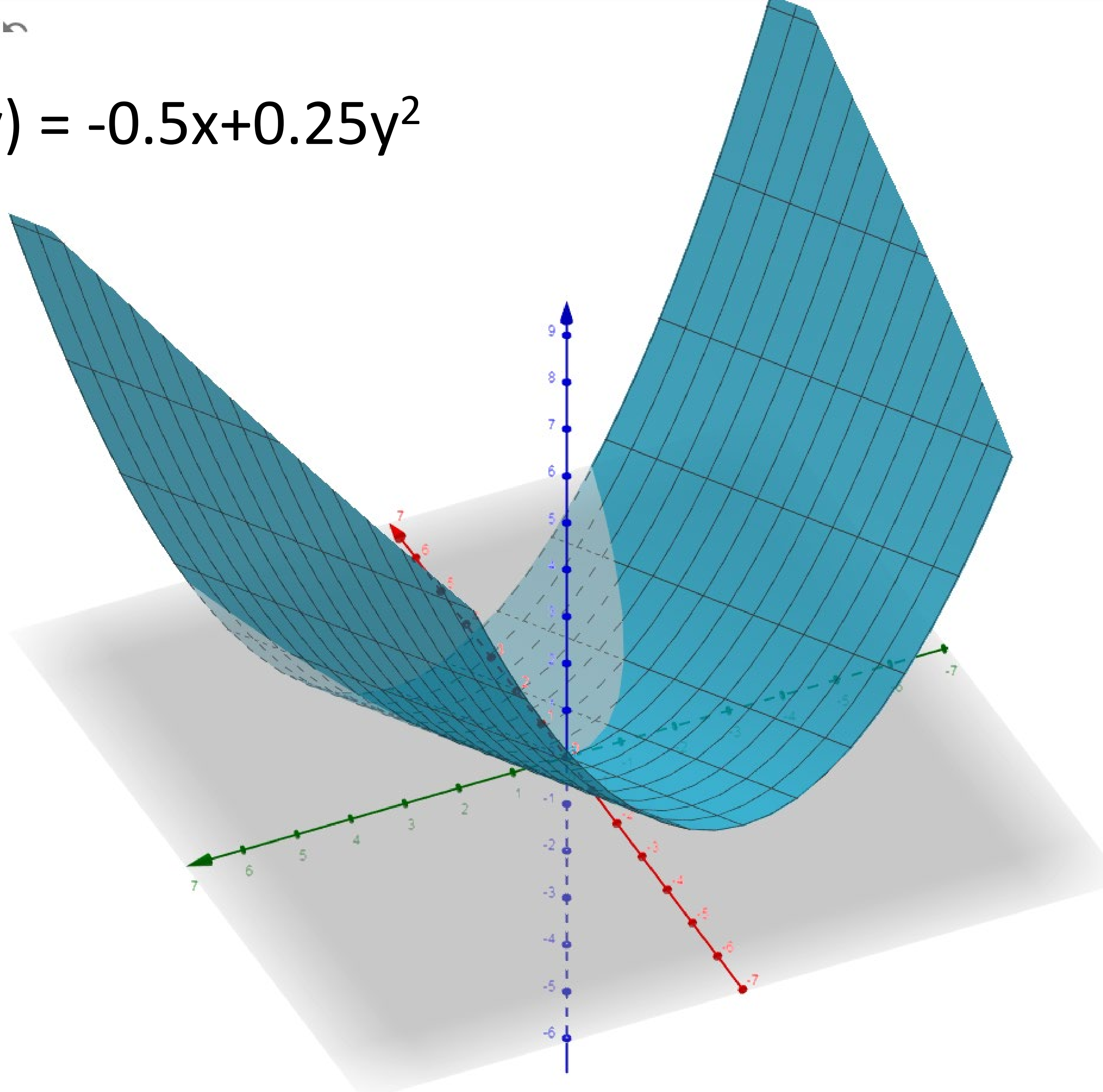


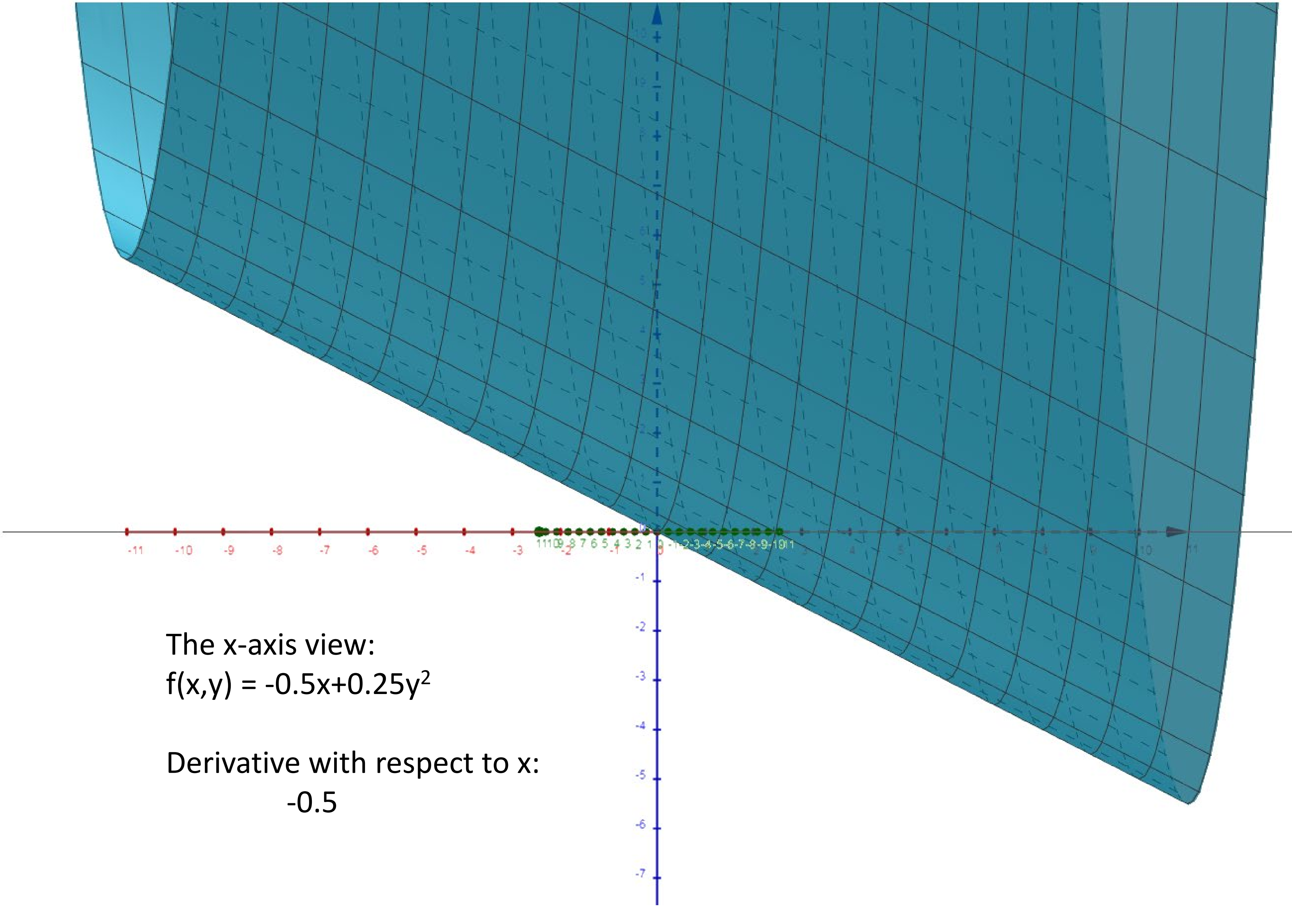
f(x,y) is a plane, so the gradient is the same vector for all points

Another example of a 3D function and its partial derivatives:

$$f(x,y) = -0.5x+0.25y^2$$

$$f(x,y) = -0.5x+0.25y^2$$





The x-axis view:

$f(x,y) = -0.5x+0.25y^2$

Derivative with respect to x:

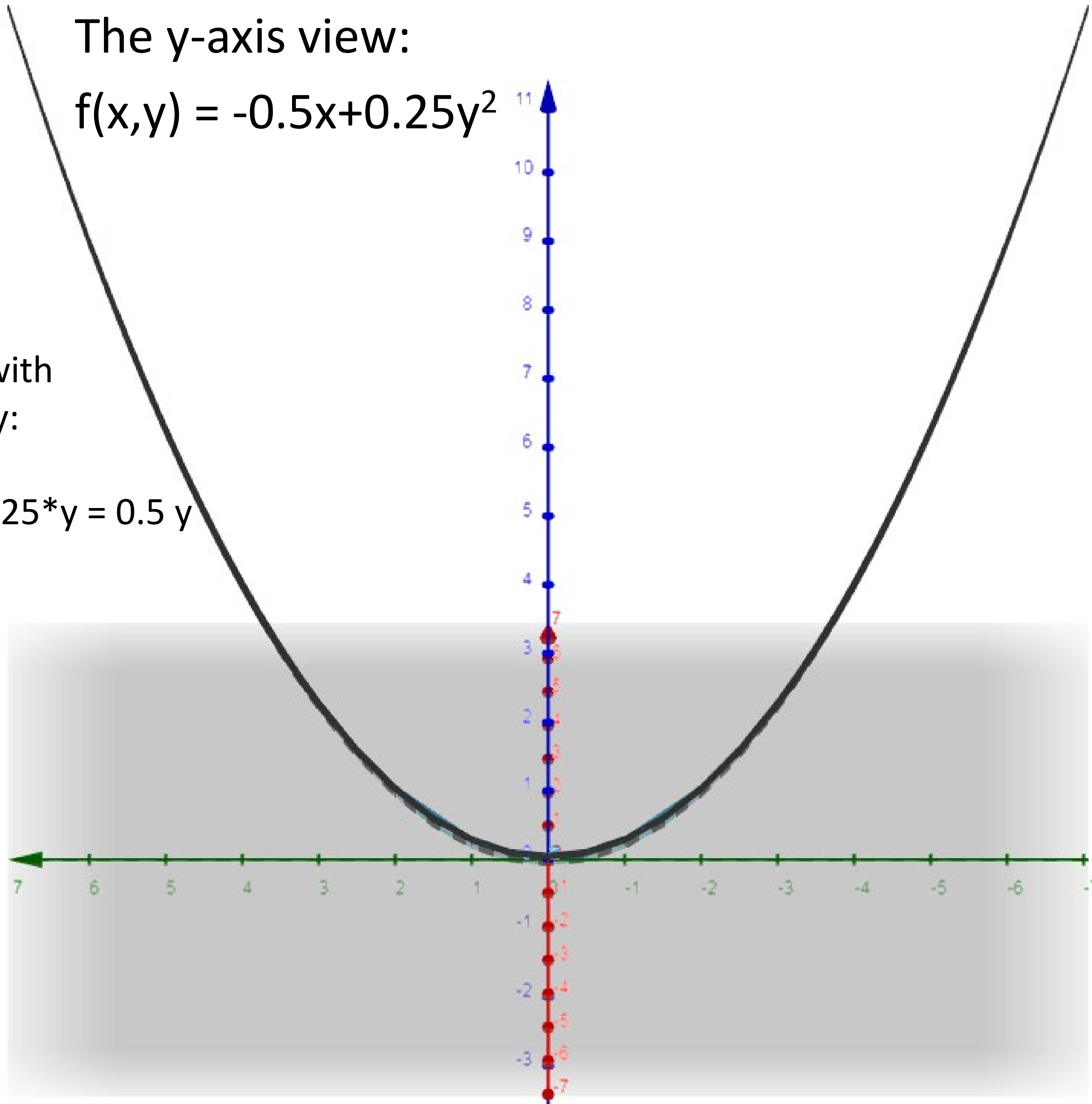
-0.5

The y-axis view:

$f(x,y) = -0.5x+0.25y^2$

Derivative with respect to y:

$= 2*0.25*y = 0.5\ y$

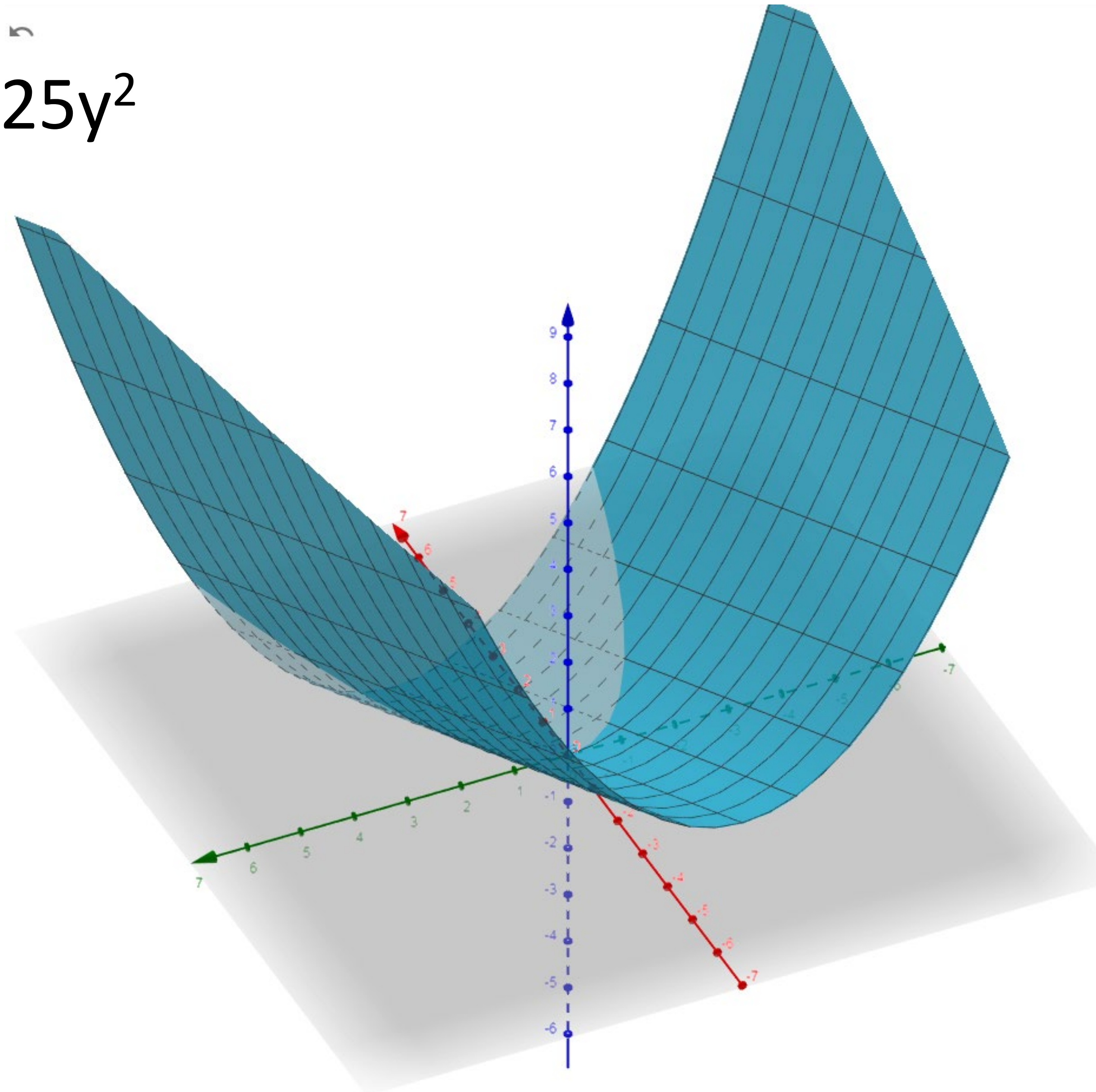


$f(x,y) = -0.5x+0.25y^2$

Gradient

∇f

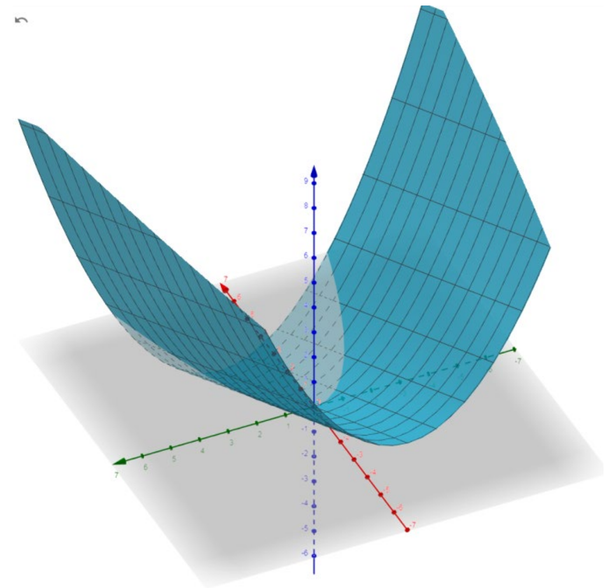
$= (-0.5,\ 0.5y)^T$



$f(x,y) = -0.5x+0.25y^2$

$= 0.5\,(-x+0.5y^2)$

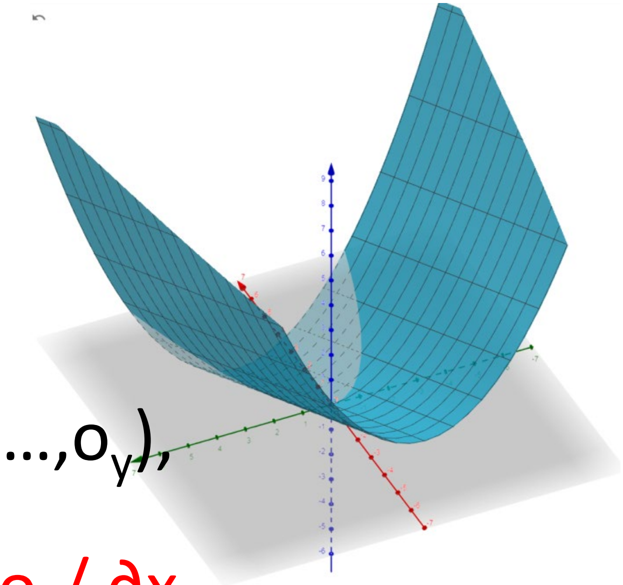
$p(o(\boldsymbol{x})) = 0.5\, o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$



$f(x,y) = -0.5x+0.25y^2$

$= 0.5\,(-x+0.5y^2)$

$p(o(\boldsymbol{x})) = 0.5\, o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$



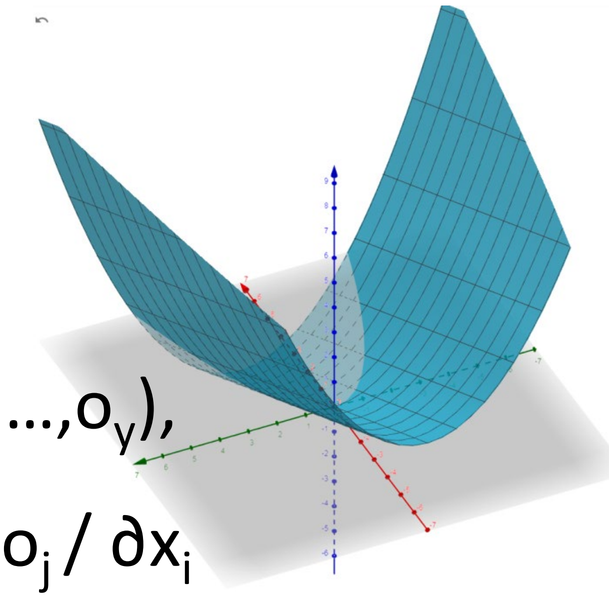
General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n),\boldsymbol{o}=(o_1,\ldots,o_y),$

$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\,\partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\,\partial o_j \;\; \partial o_j\,/\,\partial x_i$$

$f(x,y) = -0.5x+0.25y^2$

$= 0.5\,(-x+0.5y^2)$

$p(o(\boldsymbol{x})) = 0.5\, o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$



General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n)$, $\boldsymbol{o}=(o_1,\ldots,o_y)$,

$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\, \partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\, \partial o_j \;\; \partial o_j /\, \partial x_i$$

Here: J=1, f(x,y) is a scalar, n=2 => 2 partial derivatives

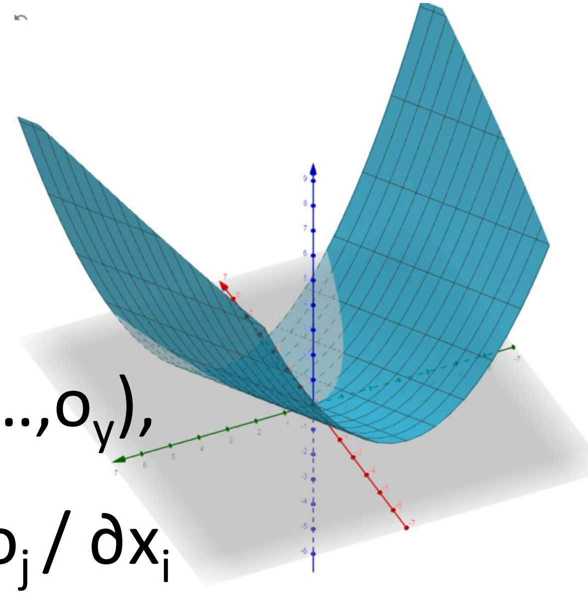
$\partial p/\partial x =$

$\partial p/\partial y =$

$f(x,y) = -0.5x+0.25y^2$

$\quad\quad = 0.5\,(-x+0.5y^2)$

$p(o(\boldsymbol{x})) = 0.5\, o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$



General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n)$, $\boldsymbol{o}=(o_1,\ldots,o_y)$,

$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\, \partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\, \partial o_j \quad \partial o_j / \partial x_i$$

Here: J=1, f(x,y) is a scalar, n=2 => 2 partial derivatives

$$\partial p(o(\boldsymbol{x}))/\, \partial x_i = \quad \partial p(o)/\, \partial o \quad \partial o / \partial x_i$$

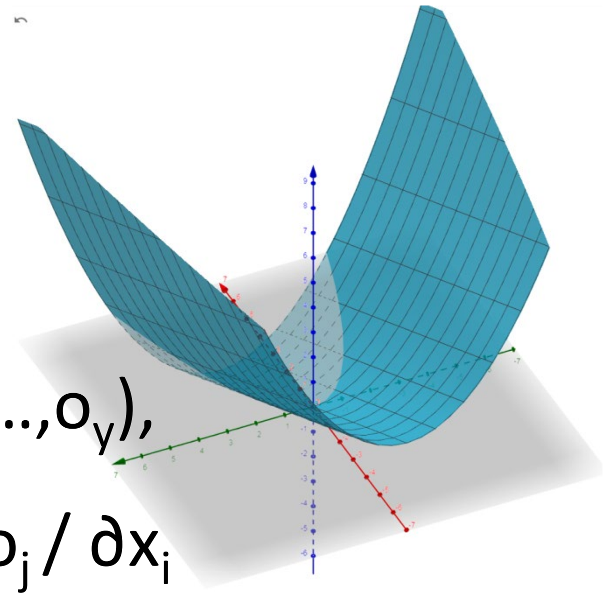
$\partial p/\partial x =$

$\partial p/\partial y =$

$f(x,y) = -0.5x+0.25y^2$

$= 0.5\,(-x+0.5y^2)$

$p(o(\boldsymbol{x})) = 0.5\,o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$



General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n),\boldsymbol{o}=(o_1,\ldots,o_y)$,

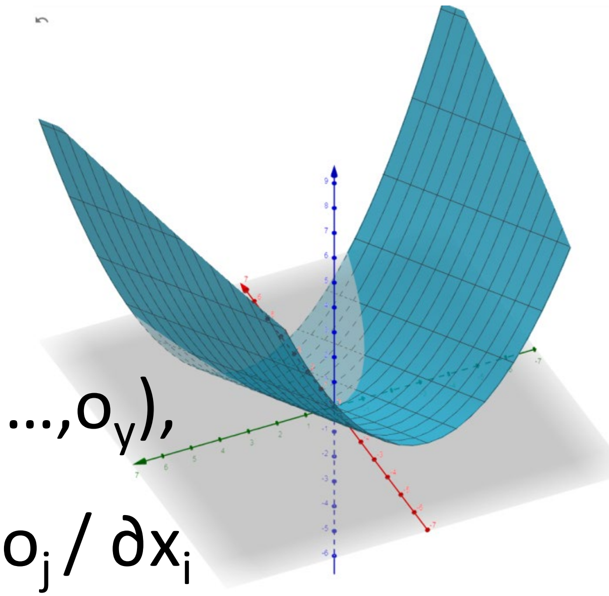
$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\partial o_j \quad \partial o_j / \partial x_i$$

Here: J=1, f(x,y) is a scalar, n=2 => 2 partial derivatives

$$\partial p(o(\boldsymbol{x}))/\partial x_i = \quad \partial p(o)/\partial o \quad \partial o / \partial x_i$$

$\partial p/\partial x =$ 0.5 -1 = -0.5

$\partial p/\partial y =$ 0.5 2 (0.5)y = 0.5 y



$f(x,y) = -0.5x+0.25y^2$

$= 0.5\,(-x+0.5y^2)$

$p(o(\boldsymbol{x})) = 0.5\, o(\boldsymbol{x})$, where $\boldsymbol{x}=(x,y)$

General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n)$, $\boldsymbol{o}=(o_1,\ldots,o_y)$,

$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\partial o_j \;\; \partial o_j/\partial x_i$$

Here: J=1, f(x,y) is a scalar, n=2 => 2 partial derivatives

$$\partial p(o(\boldsymbol{x}))/\partial x_i = \partial p(o)/\partial o \;\; \partial o/\partial x_i$$

$\partial p/\partial x = -0.5$

$\partial p/\partial y = = 0.5\, y$

$\nabla f = (-0.5, 0.5y)^T$

as computed before !