AI and Neural Network Training

Example of a 3D function and its partial derivatives:

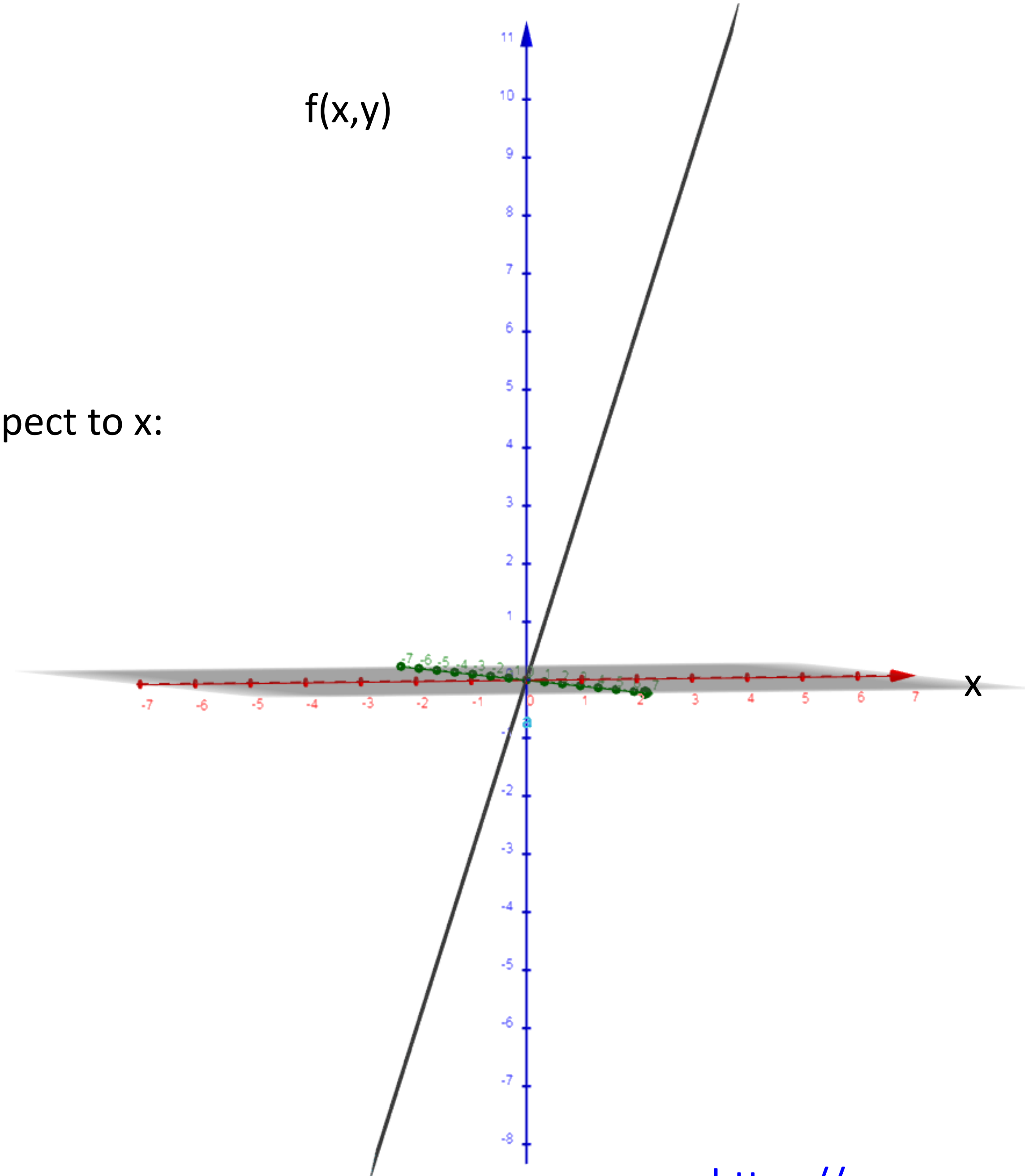
$f(x,y) = 3x+y$

f(x,y)=3x+y

Slope in x:

Derivative with respect to x:

= 3

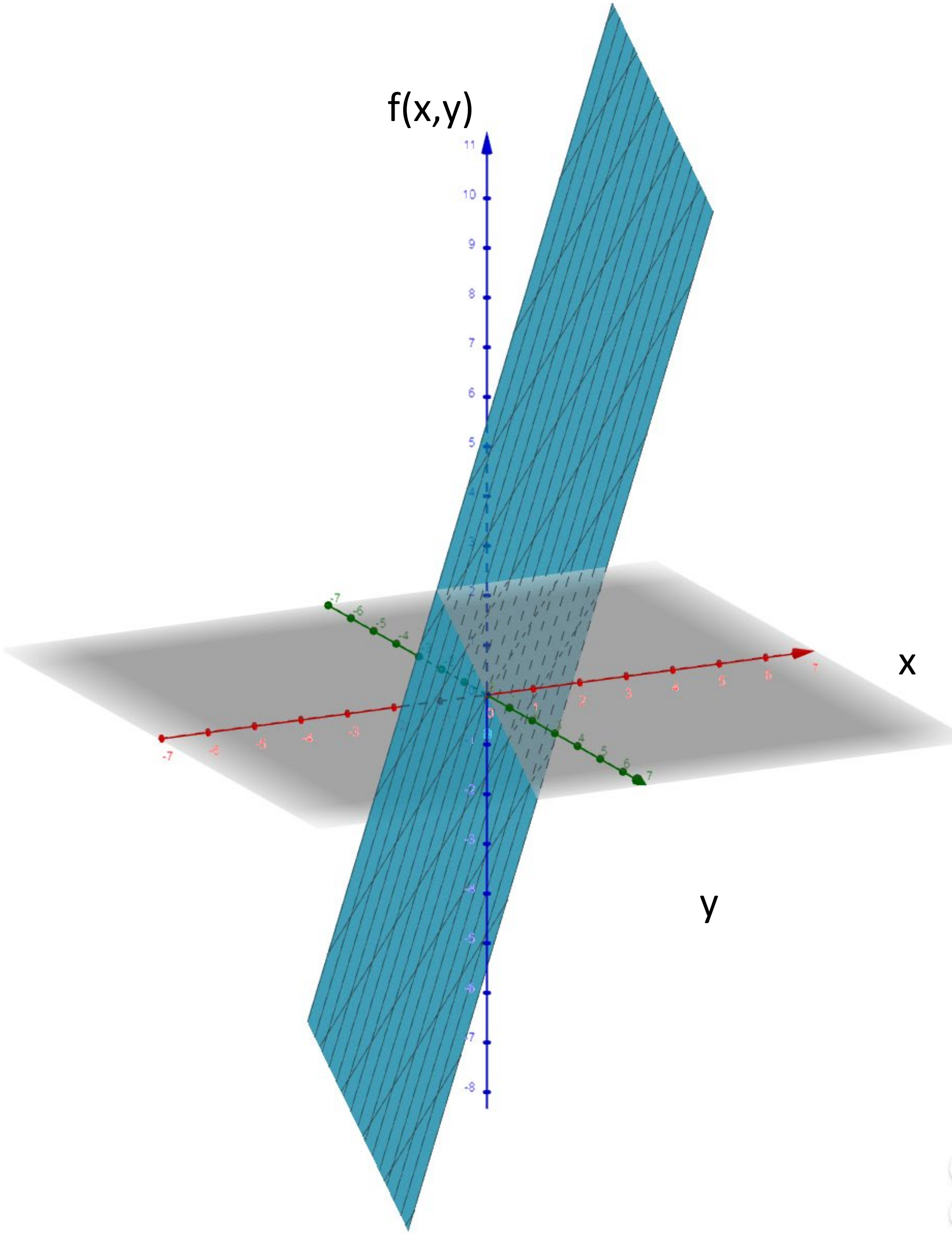


https://www.geogebra.org

f(x,y)=3x+y

Slope in y:

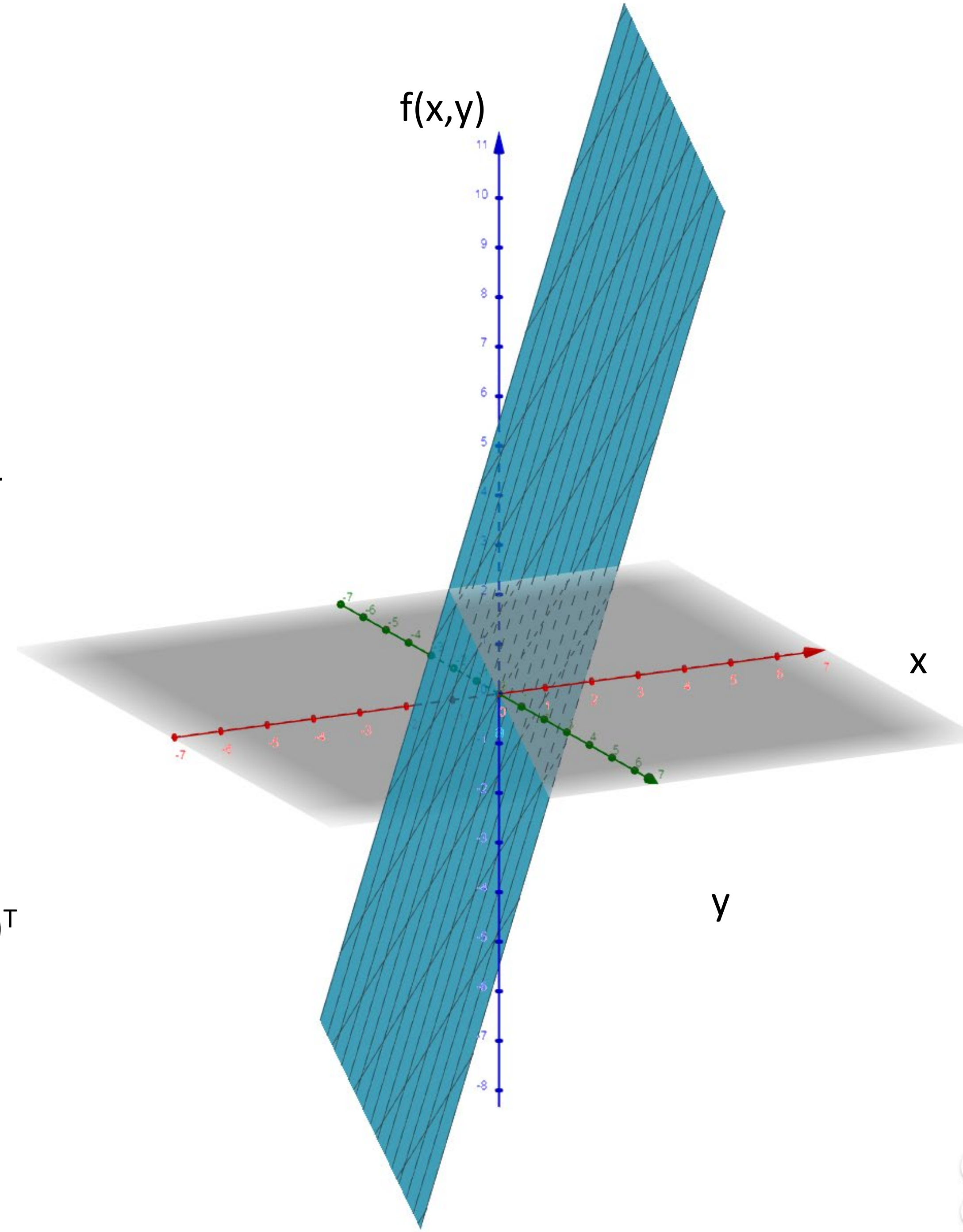
= 1



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

Slope in y:

$= 1$



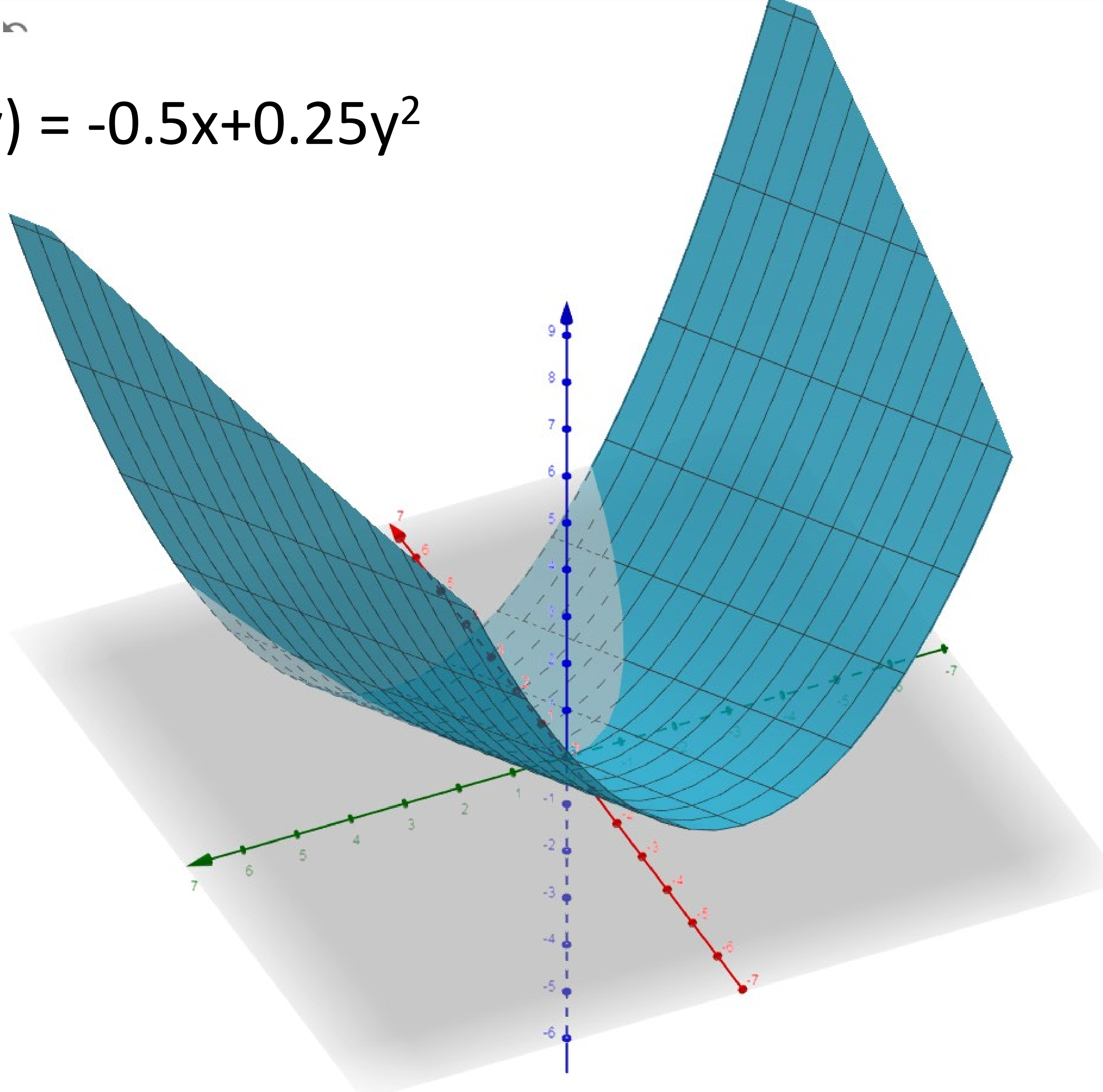
Gradient:

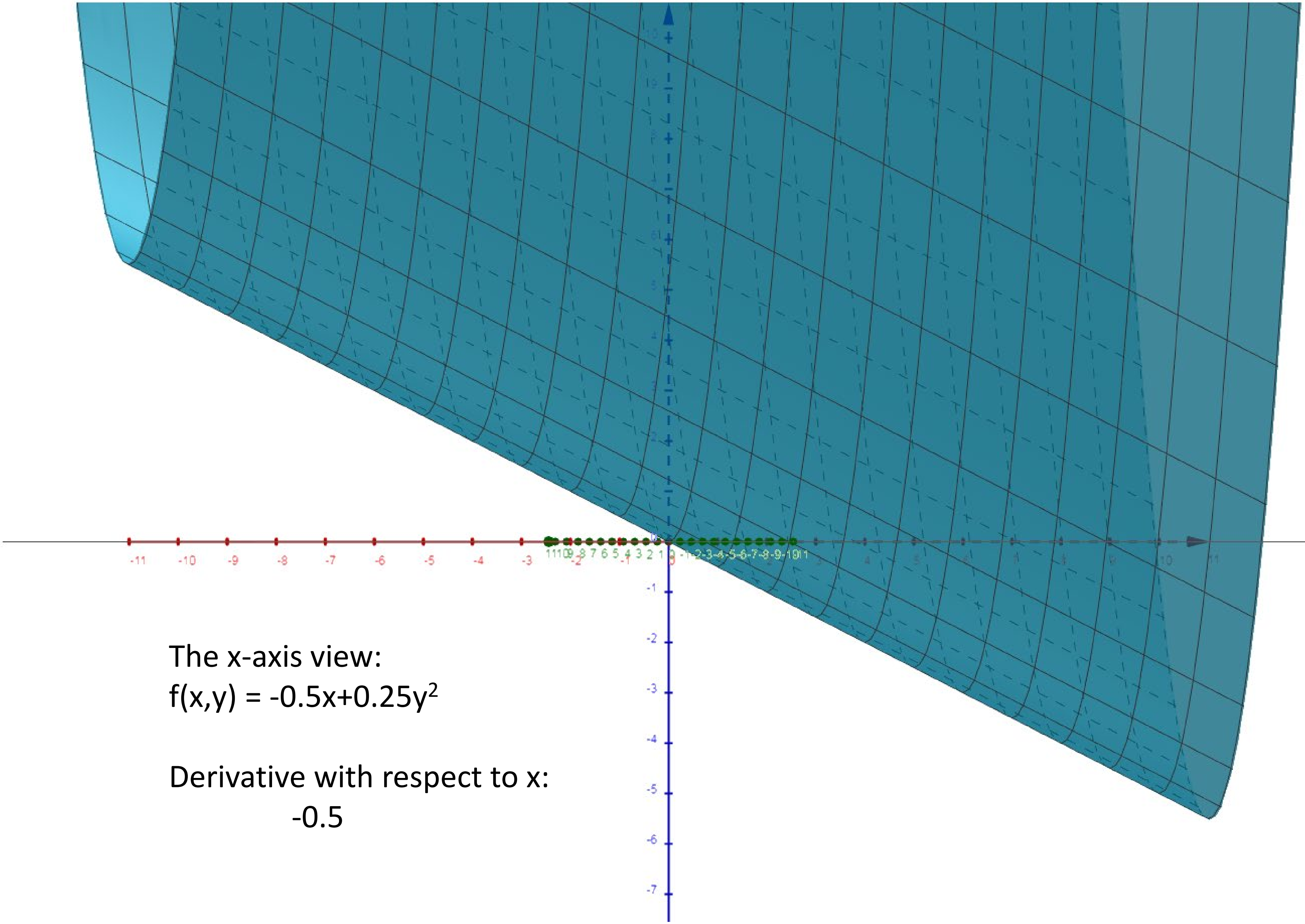
$\nabla f \quad = (3, 1)^T$

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:

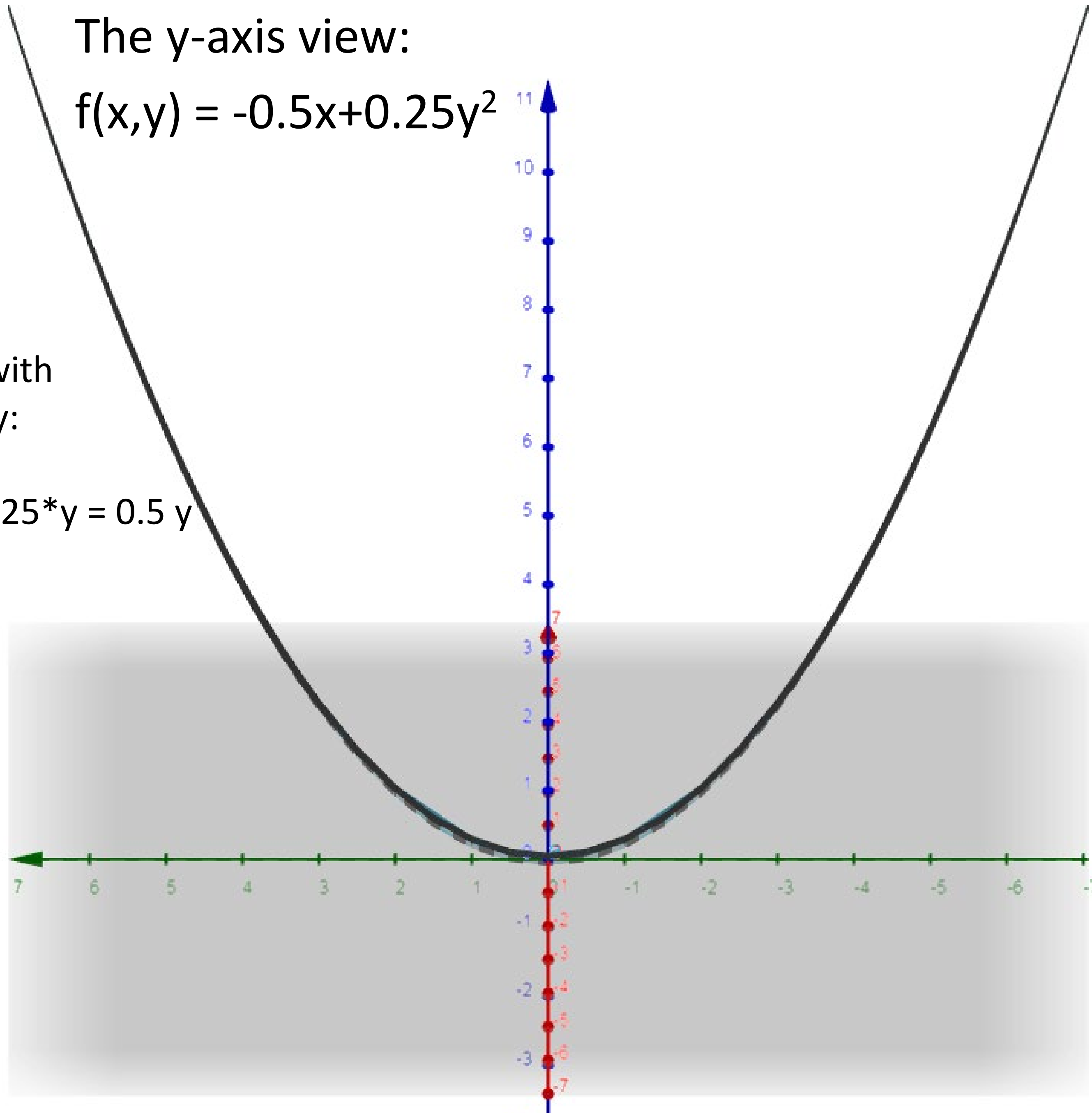
$f(x,y) = -0.5x+0.25y^2$

Derivative with respect to x:

-0.5

The y-axis view:

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



Derivative with respect to y:

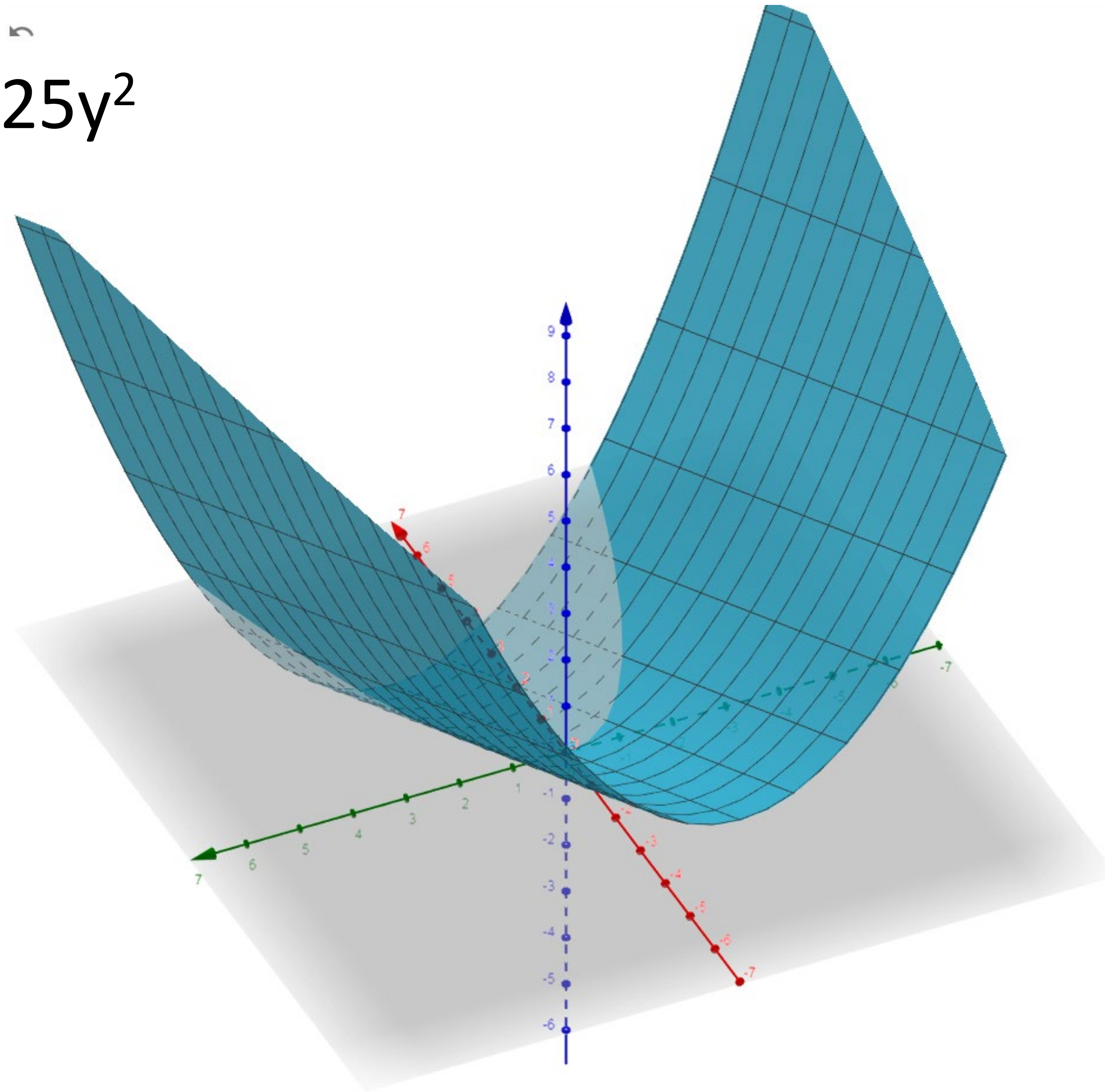
$= 2*0.25*y = 0.5\ y$

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

Gradient

∇f

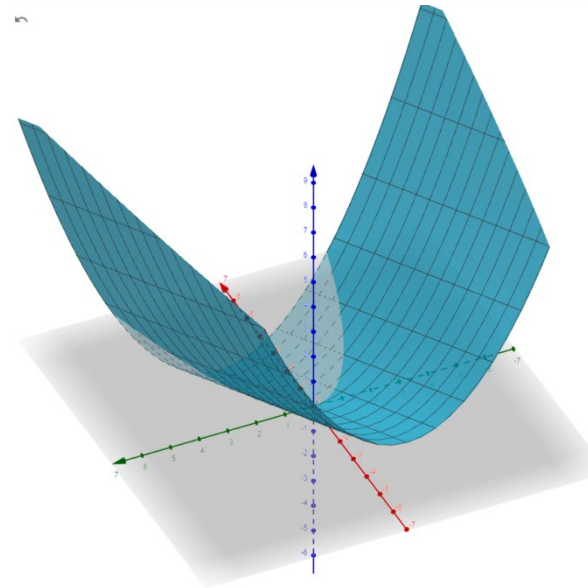
$= (-0.5, 0.5y)^T$



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

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

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



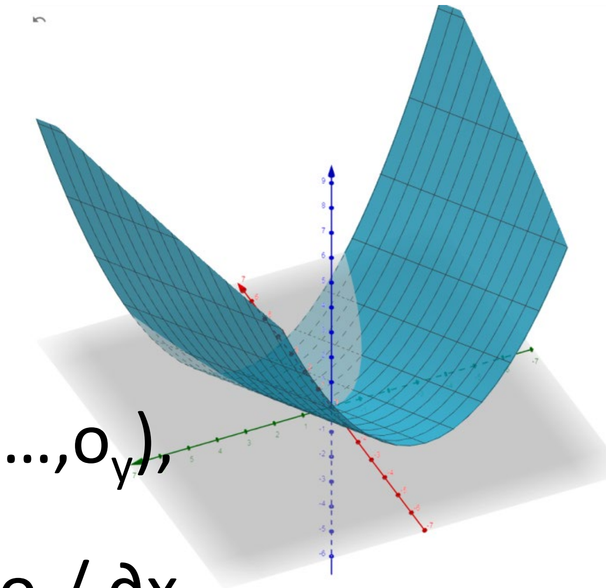
$f(x,y) = -0.5x+0.25y^2$

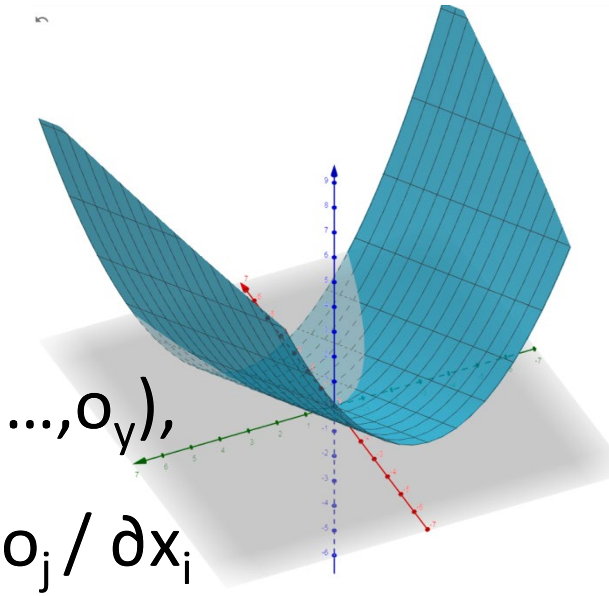
$= 0.5\,(-x+0.5y^2)$

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

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

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





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

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

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

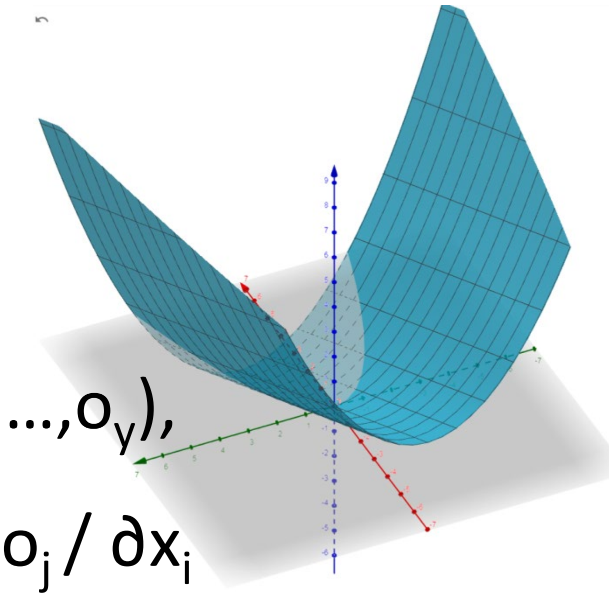
General Chain Rule: $\boldsymbol{x}=(x_1,\ldots,x_n),\boldsymbol{o}=(o_1,\ldots,o_y)$,

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

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

$\partial p/\partial x =$

$\partial p/\partial y =$



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

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

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

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

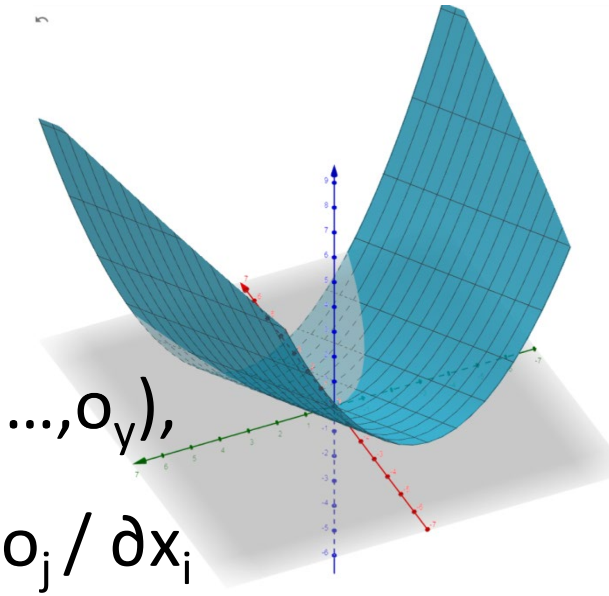
$$\partial p(\boldsymbol{o}(\boldsymbol{x}))/\,\partial x_i = \sum_{j=1,..,J} \partial p(\boldsymbol{o})/\,\partial o_j \quad \partial o_j/\,\partial x_i$$

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

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

$\partial p/\partial x =$

$\partial p/\partial y =$



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

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

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

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

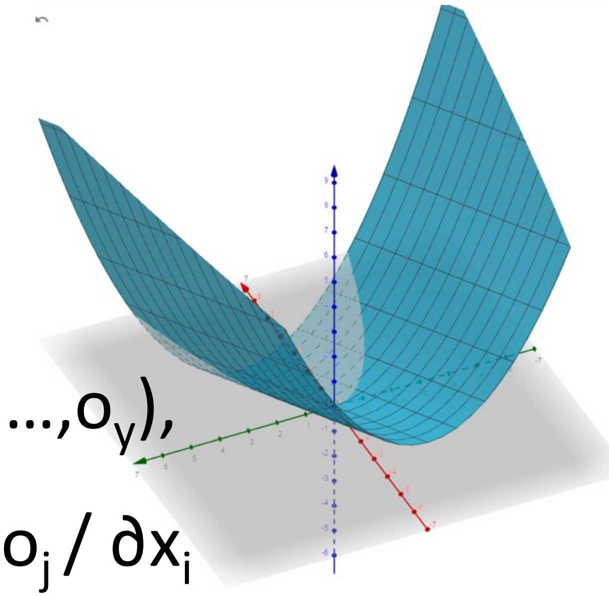
$$\partial p/\partial x = \qquad 0.5 \qquad -1 \qquad = -0.5$$

$$\partial p/\partial y = \qquad 0.5 \qquad 2\,(0.5)y \;= 0.5\, y$$

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

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

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

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

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

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

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

as computed before !