🧀 🧀 01 Gradient Descent Algorithm

Consider the function $f(x_1, x_2) = x_1^2 + 2x_2^2 + x_1x_2 + 4x_1 - 2x_2 + 5$.

We want to find the minimum. We could do it analytically, it's just a quadratic function, but quite soon we'll start to work with function for which we can't just find the optimum with pen and paper. So let's use an iterative algorithm.

0. Open up VS code, or you're favorite IDE.
1. Pick a random starting point.
2. In what direction should you move so that the function value decreases in the fastest way?
3. Move in that direction (you may consider first multiplying that direction by some small value, let's say 0.1, that's the so called α'learning rate' / 'step size', which we'll learn about later, but it basically controls how big steps you take. Too big step size -> you may overshoot the optimum value and diverge, too small, you may take too long time to converge, but often **"Let it be late, let it be almond"** principle holds)
4. Keep on iterating like that. Can you come up with a stopping criteria? (e. g. if improvement / change smaller then x let's just stop the algorithm)
5. Plot and print interesting staff, e. g. function value vs iteration number
6. Play around with α and the starting point to see how it affects convergence.

H

$$\nabla f = \begin{bmatrix} 2x_1 + x_2 + 4 \\ 4x_2 + x_1 - 2 \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$2x_1 + x_2 = -4$

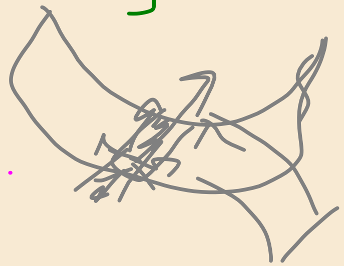
$4x_2 + x_1 = 2$

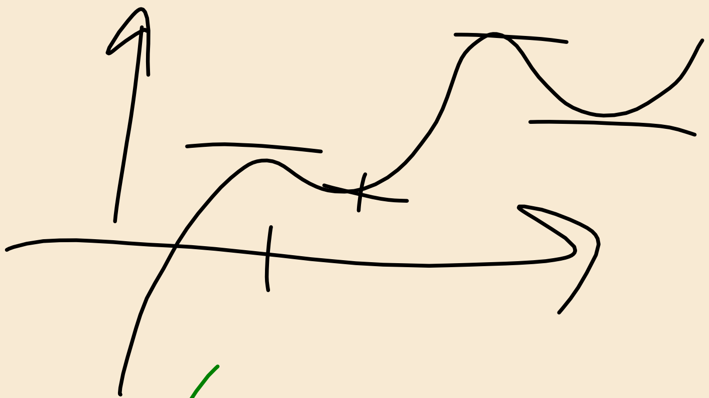
$8x_2 - x_2 = 4 - (-4)$

$\boxed{x_2 = \frac{8}{7}}$

$|2| > 0$

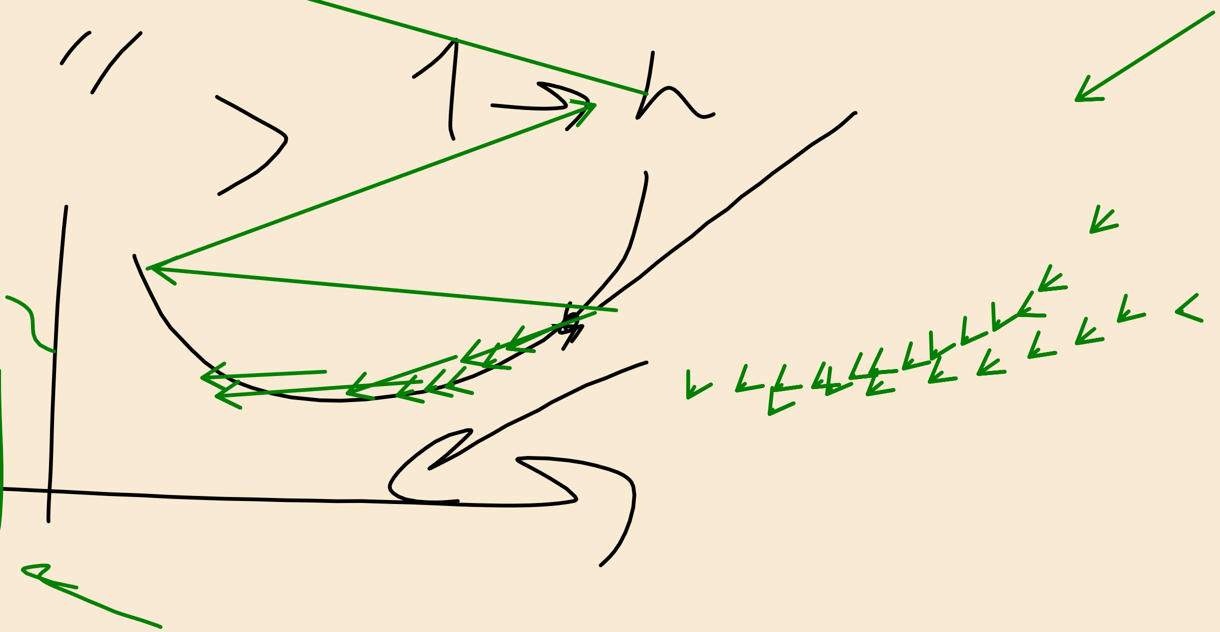
$8 - 1 > 0$





$x_1, \dots, y$

$\sum_{i=1} (y_i - f(x_i))$



🧀 02 Boat in Sevan

Սևանա լճի (x, y) կոորդինատներով կետում ջրի խորությունը

$$f(x,y) = xy^2 - 6x^2 - 3y^2$$

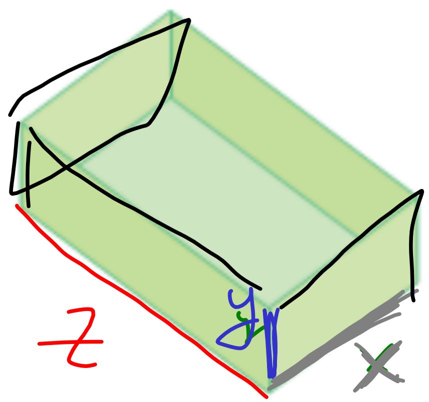
մետր է: $(5, 3)$ կետում գտնվող «Նորատուս» առագաստանավի նավապետը ցանկանում է շարժվել դեպի մի այնպիսի կետ, որում ջուրն ավելի խոր է: Նրա առաջին օգնականն առաջարկում է նավարկել դեպի հյուսիս, իսկ երկրորդը՝ դեպի հարավ: Օգնականներից որի՞ խորհուրդը պետք է լսի նավապետը:

0 $\quad 3y^2 - 12x$

$2xy$

🧀 🧀 03 Topless box

Խնդիր 7.7 Ունենք 12 մ^2 մակերեսով ստվարաթուղթ (ինչպես նաև մկրատ ու սոսինձ), որով այս անգամ ցանկանում ենք պատրաստել այսպիսի բաց տուփ (այսինքն առանց վերևի նիստի)՝



$$12 = 2xy + 2yz + xz$$

$$z(2y + x) = 12 - 2xy$$

$$z = \frac{12 - 2xy}{2y + x}$$

Ամենաշատը որքա՞ն կարող է լինել այդ տուփի ծավալը:

$$V = xy\,\frac{12 - 2xy}{2y + x}$$

Context: Smoothness

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **smooth** (or C^∞) if it has continuous partial derivatives of all orders. In practice, we often work with C^1 functions (continuously differentiable) or C^2 functions (twice continuously differentiable).

Consider the bivariate function $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto x_1^2 + 0.5x_2^2 + x_1 x_2$.

1. Show that f is smooth (continuously differentiable).
2. Find the direction of greatest increase of f at $\mathbf{x} = (1, 1)$.
3. Find the direction of greatest decrease of f at $\mathbf{x} = (1, 1)$.
4. Find a direction in which f does not instantly change at $\mathbf{x} = (1, 1)$.
5. Assume there exists a differentiable parametrization of a curve $\tilde{\mathbf{x}} : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \tilde{\mathbf{x}}(t)$ such that $\forall t \in \mathbb{R} : f(\tilde{\mathbf{x}}(t)) = f(1, 1)$. Show that at each point of the curve $\tilde{\mathbf{x}}$ the tangent vector $\frac{\partial \tilde{\mathbf{x}}}{\partial t}$ is perpendicular to $\nabla f(\tilde{\mathbf{x}})$.
6. Interpret parts (d) and (e) geometrically.

$$\nabla f(1,1) = \begin{bmatrix} 2x_1 + x_2 \\ x_2 + x_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left\| \begin{bmatrix} v_x \\ v_y \end{bmatrix} \right\| \quad \left\| \nabla f \begin{bmatrix} v_x \\ v_y \end{bmatrix} \right\|$$

$$3v_x + 2v_y = 0$$

$$v_x = -\frac{2}{3} v_y$$

3 -2 0

2 3 0

$$f_x v_x + f_y v_y$$

1 0

0 1 0

$$\underbrace{\nabla f}_{\mathbb{R}^2 \to \mathbb{R}} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

(3,0)

(2,3)

∇f

🧀 🧀 🧀 05 Contour Plots, Hessians, and Convexity

Context: Contour Plots and Convexity

Contour plots (or level curves) visualize multivariate functions by showing curves where $f(x_1, x_2) = c$ for various constants c. They're essential for understanding the shape of loss landscapes in machine learning.

Let $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto -\cos(x_1^2 + x_2^2 + x_1 x_2)$.

1. Create a contour plot of f in the range $[-2, 2] \times [-2, 2]$ with R.
2. Compute ∇f.
3. Compute $\nabla^2 f$ (the Hessian matrix).

Now, define the restriction of f to $S_r = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 + x_1 x_2 < r\}$ with $r \in \mathbb{R}, r > 0$, i.e., $f|_{S_r} : S_r \to \mathbb{R}, (x_1, x_2) \mapsto f(x_1, x_2)$.

4. Show that $f|_{S_r}$ with $r = \pi/4$ is convex.
5. Find the local minimum $\mathbf{x}^*$ of $f|_{S_r}$.
6. Is $\mathbf{x}^*$ a global minimum of f?

🧀 🧀 06 Taylor Expansion

Consider the bivariate function
$f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto \exp(\pi \cdot x_1) - \sin(\pi \cdot x_2) + \pi \cdot x_1 \cdot x_2.$

1. Compute the gradient of f for an arbitrary x.
2. Compute the Hessian of f for an arbitrary x.
3. State the first order taylor polynomial $T_{1,a}(x)$ expanded around the point $a = (0, 1)$.
4. State the second order taylor polynomial $T_{2,a}(x)$ expanded around the point $a = (0, 1)$.
5. Determine if $T_{2,a}$ is a convex function.