Assignment 1
Optimisation for Computer Science
Group 33

Sven Celin & Gabriela Ožegović

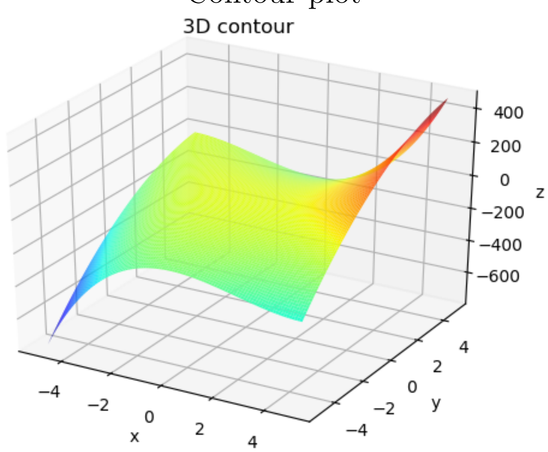
November 2019

1 Characterization of Function

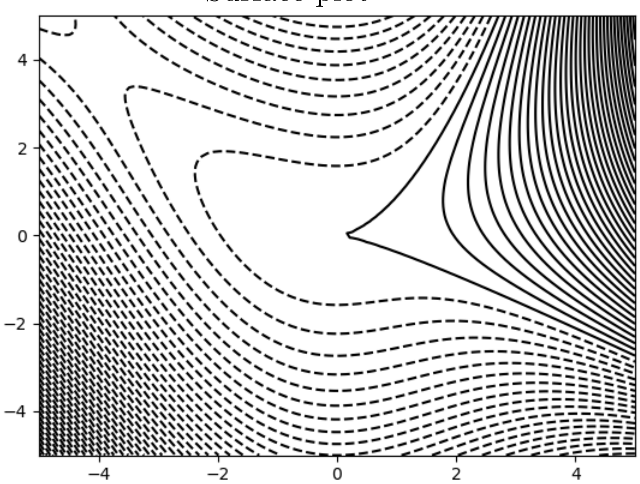
1.

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

Contour plot



Surface plot



First derivatives

$$\frac{\partial f}{\partial x} = 6x^2 + 6xy \qquad \frac{\partial f}{\partial y} = -12y + 3x^2$$

Second derivatives

$$\frac{\partial^2 f}{\partial x^2} = 12x + 6y \qquad \frac{\partial^2 f}{\partial x \partial y} = 6x$$

$$\frac{\partial^2 f}{\partial y \partial x} = 6x \qquad \frac{\partial^2 f}{\partial y^2} = -12$$

Gradient (vector of first partial derivatives):

$$\begin{bmatrix} 6x^2 + 6xy \\ -12y + 3x^2 \end{bmatrix}$$

Hessian matrix (matrix of second partial derivatives):

$$\begin{bmatrix} 12x + 6y & 6x \\ 6x & -12 \end{bmatrix}$$

Since the rate of change of a function is zero at a stationary point, it is necessary to equalize the first derivations with zero, and then calculate the obtained system of equations with two unknowns.

$$\begin{aligned} 6x^2 + 6xy &= 0 \\ -12y + 3x^2 &= 0 \end{aligned}$$

After dividing with a common divisor and excretion where possible we get:

$$\begin{aligned} x(x+y) &= 0 \\ x^2 - 4y &= 0 \end{aligned}$$

From the first equation we get x.

$$x(x+y) = 0 \quad \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = -y \end{cases}$$

We use calculated x to get y pairs from the second equation.

$$y^2 - 4y = 0$$

$$y(y-4) = 0 \quad \Rightarrow \begin{cases} y_1 = 0 \\ y_2 = 4 \Rightarrow x_2 = -4 \end{cases}$$

Our stationary points are:

$$P_1(0,0) \text{ and } P_2(-4,4)$$

In order to use the second derivative test to evaluate these stationary points to see whether or not either of them is a maximum, minimum or a saddle point we need to include the x and y of each point in second order partial derivatives.

$$\frac{\partial^2 f}{\partial x^2} = 12x + 6y \quad \Rightarrow \begin{cases} P_1(0,0) = 0 \\ P_2(-4,4) = -24 \end{cases}$$

$$\frac{\partial^2 f}{\partial y \partial x} = 6x \quad \Rightarrow \begin{cases} P_1(0,0) = 0 \\ P_2(-4,4) = -24 \end{cases}$$

$$\frac{\partial^2 f}{\partial y^2} = -12 \quad \Rightarrow \begin{cases} P_1(0,0) = -12 \\ P_2(-4,4) = -12 \end{cases}$$

To evaluate each point we need to calculate the determinant of the Hessian matrix for each one, using the calculations above.
The equation for the determinant of the Hessian matrix is:

$$D(M) = \frac{\partial^2 f}{\partial x^2} * \frac{\partial^2 f}{\partial y^2} - (\frac{\partial^2 f}{\partial y \partial x})^2$$

Determinants for our points are:

$$D(M)_1 = 0 * (-12) - 0^2 = 0$$

$$D(M)_2 = -24 * (-12) - (-24)^2 = -288$$

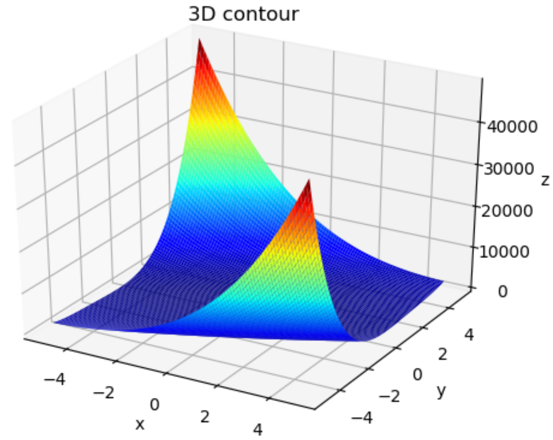
If the determinant is equal to zero, the point need further examination.
If the determinant is less than zero, the point is a saddle point.
If the determinant is greater than zero, we need to check the value of the point in $\frac{\partial^2 f}{\partial x^2}$.
If that value is less than zero, the point is a maximum. If it is greater than zero, the point is a minimum.

Point $P_1(0,0)$ needs further examination. By looking at graph, we conclude that it is a saddle point.
Point $P_1(-4,4)$ is a saddle point.

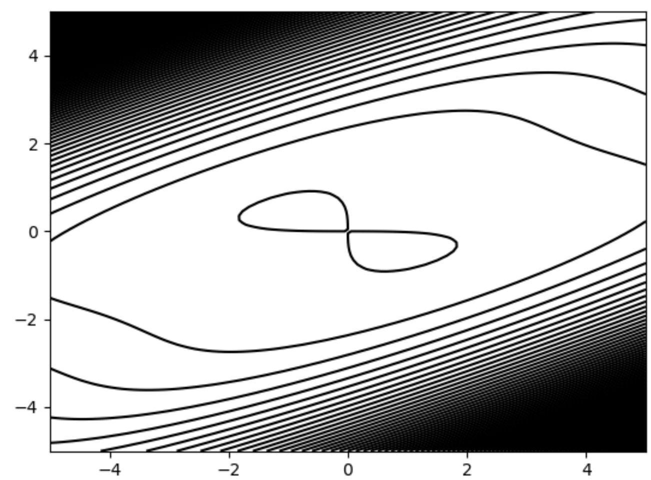
2.

$$f(x,y) = (x-2y)^4 + 64xy$$

Contour plot



Surface plot



First derivatives

$$\frac{\partial f}{\partial x} = 4(x-2y)^3 + 64y \qquad \frac{\partial f}{\partial y} = -8(x-2y)^3 + 64x$$

Second derivatives

$$\frac{\partial^2 f}{\partial x^2} = 12(x-2y)^2 \qquad \frac{\partial^2 f}{\partial x \partial y} = -24(x-2y)^2 + 64$$
$$\frac{\partial^2 f}{\partial y \partial x} = -24(x-2y)^2 + 64 \qquad \frac{\partial^2 f}{\partial y^2} = 48(x-2y)^2$$

Gradient:

$$\begin{bmatrix} 4(x-2y)^3 + 64y \\ -8(x-2y)^3 + 64x \end{bmatrix}$$

Hessian matrix:

$$\begin{bmatrix} 12(x-2y)^2 & -24(x-2y)^2 + 64 \\ -24(x-2y)^2 + 64 & 48(x-2y)^2 \end{bmatrix}$$

System of equations with two unknowns:

$$
\begin{aligned}
4(x-2y)^3 + 64y &= 0 \\
-8(x-2y)^3 + 64x &= 0
\end{aligned}
$$

$$
\begin{aligned}
\cancel{(x-2y)^3} + 16y &= 0 \\
\cancel{(x-2y)^3} - 8x &= 0 \quad | \quad -
\end{aligned}
$$

$$
\begin{aligned}
16y + 8x &= 0 \\
x &= -2y
\end{aligned}
$$

We insert the x in first equation.

$$
\begin{aligned}
(-4y)^3 + 16y &= 0 \\
-64y^3 + 18y &= 0 \quad | \quad / -16 \\
4y^3 - y &= 0 \\
4(4y^2 - 1) &= 0 \\
y(2y-1)(2y-1) &= 0
\end{aligned}
$$

$$
y(2y-1)(2y-1) = 0 \quad \Rightarrow \begin{cases} y_1 = 0 \Rightarrow x_1 = 0 \\ y_2 = \dfrac{1}{2} \Rightarrow x_2 = -1 \\ y_2 = -\dfrac{1}{2} \Rightarrow x_3 = 1 \end{cases}
$$

Our stationary points are:

$$P_1(0,0),\ P_2(\tfrac{1}{2}, -1) \text{ and } P_3(-\tfrac{1}{2}, 1)$$

Second derivative test:

$$\frac{\partial^2 f}{\partial x^2} = 12(x-2y)^2 \quad \Rightarrow \begin{cases} P_1(0,0) = 0 \\ P_2(\frac{1}{2},-1) = 30 \\ P_3(-\frac{1}{2},1) = -30 \end{cases}$$

$$\frac{\partial^2 f}{\partial y \partial x} = -24(x-2y)^2 + 64 \quad \Rightarrow \begin{cases} P_1(0,0) = 64 \\ P_2(\frac{1}{2},-1) = -86 \\ P_3(-\frac{1}{2},1) = -86 \end{cases}$$

$$\frac{\partial^2 f}{\partial y^2} = 48(x-2y)^2 \quad \Rightarrow \begin{cases} P_1(0,0) = 0 \\ P_2(\frac{1}{2},-1) = -300 \\ P_3(-\frac{1}{2},1) = -300 \end{cases}$$

Determinants for our points are:

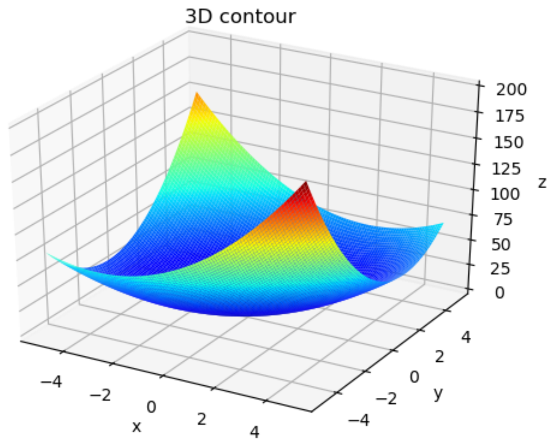
$$D(M)_1 = -4096$$
$$D(M)_2 = 15104$$
$$D(M)_3 = 15104$$

Point $P_1(0,0)$ is a saddle point.
Point $P_2(\frac{1}{2},-1)$ is a local minimum ($\frac{\partial^2 f}{\partial x^2} > 0$).
Point $P_3(-\frac{1}{2},1)$ is a local maximum ($\frac{\partial^2 f}{\partial x^2} < 0$).

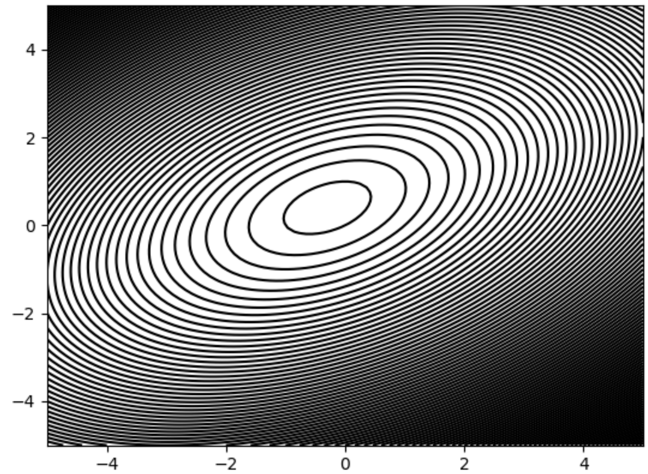
3.

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

Contour plot



Surface plot



First derivatives

$$\frac{\partial f}{\partial x} = 4x - 2y + 2 \qquad \frac{\partial f}{\partial y} = 6y - 2x + 3$$

Second derivatives

$$\frac{\partial^2 f}{\partial x^2} = 4 \qquad \frac{\partial^2 f}{\partial x \partial y} = -2$$

$$\frac{\partial^2 f}{\partial y \partial x} = -2 \qquad \frac{\partial^2 f}{\partial y^2} = 6$$

Gradient:

$$\begin{bmatrix} 4x - 2y + 2 \\ 6y - 2x + 3 \end{bmatrix}$$

Hessian matrix:

$$\begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix}$$

System of equations with two unknowns:

$$\begin{aligned} 4x - 2y + 2 &= 0 \\ -2x + 6y - 3 &= 0 \end{aligned}$$

$$\begin{aligned} \cancel{2x} - y + 1 &= 0 \\ \cancel{-2x} + 6y - 3 &= 0 \end{aligned}$$

$$\begin{aligned} 5y - 2 &= 0 \\ y &= \frac{2}{5} \end{aligned}$$

We insert the y in first equation.

$$\begin{aligned} 4x - 2 * \frac{2}{5} + 2 &= 0 \\ 4x + \frac{6}{5} &= 0 \\ x &= -\frac{3}{10} \end{aligned}$$

Our stationary point is:

$$P_1(-\tfrac{3}{10}, \tfrac{2}{5})$$

Second derivative test:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= 4 \quad \Rightarrow \left\{ P_1(-\frac{3}{10}, \frac{2}{5}) = 4 \right. \\ \frac{\partial^2 f}{\partial y \partial x} &= -2 \quad \Rightarrow \left\{ P_1(-\frac{3}{10}, \frac{2}{5}) = -2 \right. \\ \frac{\partial^2 f}{\partial y^2} &= 6 \quad \Rightarrow \left\{ P_1(-\frac{3}{10}, \frac{2}{5}) = 6 \right. \end{aligned}$$

Determinants for our point is:

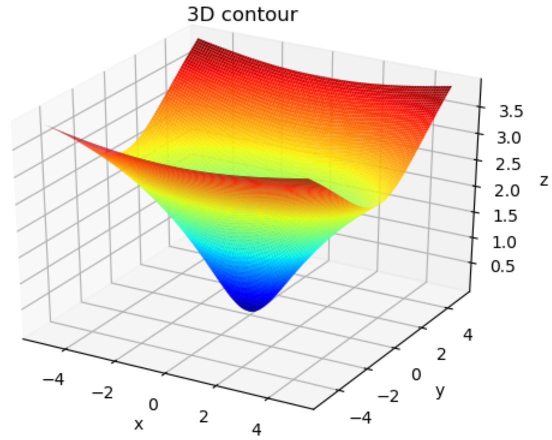
$$D(M)_1 = 20$$

Point $P_1(-\frac{3}{10}, \frac{2}{5})$ is a local minimum ($\frac{\partial^2 f}{\partial x^2} > 0$).

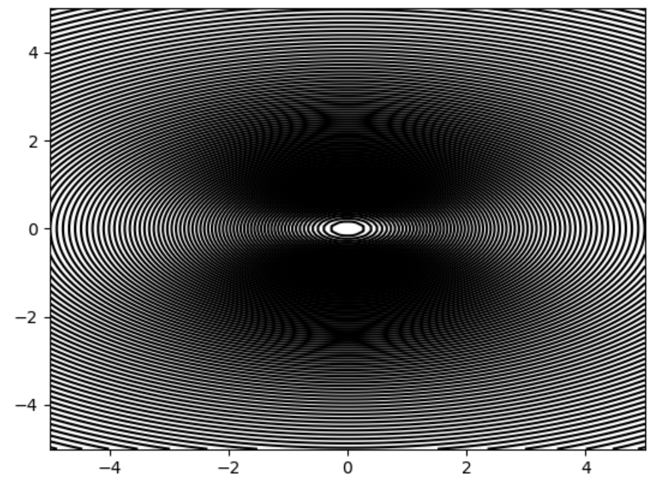
4.

$$f(x,y) = \ln(1 + \frac{1}{2}(x^2 + 3y^2))$$

Contour plot



Surface plot



First derivatives

$$\frac{\partial f}{\partial x} = \frac{2x}{2 + x^2 + 3y^2} \qquad \frac{\partial f}{\partial y} = \frac{6y}{2 + x^2 + 3y^2}$$

Second derivatives

$$\frac{\partial^2 f}{\partial x^2} = \frac{4 + 2x^2 + 6y^2 + 4x^2}{(2 + x^2 + 3y^2)^2} \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{-2x * 6y}{(2 + x^2 + 3y^2)^2}$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{-2x * 6y}{(2 + x^2 + 3y^2)^2} \qquad \frac{\partial^2 f}{\partial y^2} = \frac{12 + 6x^2 - 18y^2}{(2 + x^2 + 3y^2)^2}$$

Gradient:

$$\begin{bmatrix} \frac{2x}{2+x^2+3y^2} \\ \frac{6y}{2+x^2+3y^2} \end{bmatrix}$$

Hessian matrix:

$$\begin{bmatrix} \frac{4+2x^2+6y^2+4x^2}{(2+x^2+3y^2)^2} & \frac{-2x*6y}{(2+x^2+3y^2)^2} \\ \frac{-2x*6y}{(2+x^2+3y^2)^2} & \frac{12+6x^2-18y^2}{(2+x^2+3y^2)^2} \end{bmatrix}$$

System of equations with two unknowns:

$$\frac{2x}{2+x^2+3y^2} = 0 \qquad \Rightarrow \begin{cases} 2x = 0 \\ x = 0 \end{cases}$$

$$\frac{6y}{2+x^2+3y^2} = 0 \qquad \Rightarrow \begin{cases} 6y = 0 \\ y = 0 \end{cases}$$

Our stationary point is:

$$P_1(0,0)$$

Second derivative test:

$$\frac{\partial^2 f}{\partial x^2} = \frac{4+2x^2+6y^2+4x^2}{(2+x^2+3y^2)^2} \quad \Rightarrow \{P_1(0,0) = 1$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{-2x*6y}{(2+x^2+3y^2)^2} \quad \Rightarrow \{P_1(0,0) = 0$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{12+6x^2-18y^2}{(2+x^2+3y^2)^2} \quad \Rightarrow \{P_1(0,0) = 3$$

Determinants for our point is:

$$D(M)_1 = 3$$

Point $P_1(0,0)$ is a local minimum ($\frac{\partial^2 f}{\partial x^2} > 0$).

2 Numerical Gradient Approximation

We are analytically computing gradient by choosing one random point, $P(0.5, 0.4)$, and calculating the value of first partial derivatives (elements of gradient vector) in that point. In python script, called *functions.py*, we are computing a numerical approximation using central differences in the same point.

1.
$$f(x, y) = 2x^3 - 6y^2 + 3x^2y$$

$$\begin{aligned}\frac{\partial f}{\partial x} = 6^2 + 6xy &= 6 * (0.5)^2 + 6 * 0.5 * 0.4 \\ &= 2.7 \\ \frac{\partial f}{\partial y} = -12y + 3x^2 &= -12 * (0.4) + 3 * (0.5)^2 \\ &= -4.05\end{aligned}$$

Values calculated with python script are:

$$\nabla_x f(x, y) \approx \frac{f(x + \epsilon, y) - f(x - \epsilon, y)}{2\epsilon} = 2.700001999999979$$

$$\nabla_y f(x, y) \approx \frac{f(x, y + \epsilon) - f(x, y - \epsilon)}{2\epsilon} = -4.0500000000000815$$

2.
$$f(x, y) = (x - 2y)^4 + 64xy$$

$$\begin{aligned}\frac{\partial f}{\partial x} = 4(x - 2y)^3 + 64y &= 4(0.5 - 2 * 0.4)^3 + 64 * 0.4 \\ &= 25.492 \\ \frac{\partial f}{\partial y} = -8(x - 2y)^3 + 64x &= -8(0.5 - 2 * 0.4)^3 + 64 * 0.5 \\ &= 32.216\end{aligned}$$

Values calculated with python script are:

$$\nabla_x f(x,y) \approx \frac{f(x+\epsilon,y) - f(x-\epsilon,y)}{2\epsilon} = 25.491998800000637$$

$$\nabla_y f(x,y) \approx \frac{f(x,y+\epsilon) - f(x,y-\epsilon)}{2\epsilon} = 32.21600959999993$$

3.

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

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 4x - 2y + 2 = 4 * 0.5 - 2 * 0.4 + 2 \\
&= 3.2 \\
\frac{\partial f}{\partial y} &= 6y - 2x + 3 = 6 * 0.4 - 2 * 0.5 + 3 \\
&= -1.6
\end{aligned}$$

Values calculated with python script are:

$$\nabla_x f(x,y) \approx \frac{f(x+\epsilon,y) - f(x-\epsilon,y)}{2\epsilon} = 3.1999999999999806$$

$$\nabla_y f(x,y) \approx \frac{f(x,y+\epsilon) - f(x,y-\epsilon)}{2\epsilon} = -1.5999999999999348$$

4.

$$f(x,y) = \ln(1 + \frac{1}{2}(x^2 + 3y^2))$$

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{2x}{2 + x^2 + 3y^2} = \frac{2 * 0.5}{2 + 0.5^2 + (3 * 0.4)^2} \\
&= 0.366 \\
\frac{\partial f}{\partial y} &= \frac{6y}{2 + x^2 + 3y^2} = \frac{6 * 0.4}{2 + 0.5^2 + (3 * 0.4)^2} \\
&= 0.879
\end{aligned}$$

Values calculated with python script are:

$$\nabla_x f(x,y) \approx \frac{f(x+\epsilon,y)-f(x-\epsilon,y)}{2\epsilon} = 0.3663002485073119$$

$$\nabla_y f(x,y) \approx \frac{f(x,y+\epsilon)-f(x,y-\epsilon)}{2\epsilon} = 0.8791201395315917$$

3 Vectors, Norms and Matrices

1. Norm $\| \cdot \|_{\frac{1}{2}}$ not a norm because of the triangle inequality. Triangle inequality is true if $\| \cdot \|_n \ for \ n < 1$.

   Let n < 1 then:

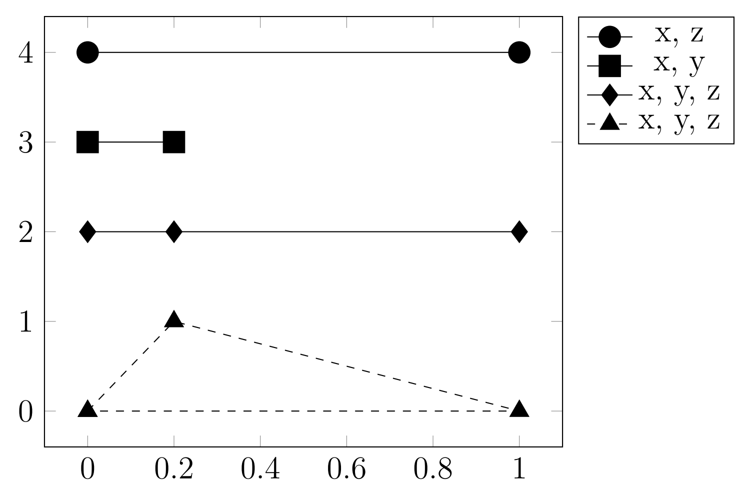
$$\|(1,0,0,...,0...)\|_n = 1$$
$$\underline{\|(0,1,0,...,0...)\|_n = 1/+ \qquad}$$
$$\|(1,1,0,...,0...)\|_n = 2^{1/n} > 2$$

2. In next figure we will explain why does $\|x-z\| \leq \|x-y\| + \|y-z\|$ for $x, y, z \in \mathbb{R}^n$ hold using triangle equality.

$$\|a-b\| \leq \|a\| + \|b\|$$
$$a = x - y$$
$$b = y - z$$

then we replace norm values with a and b

$$\|x-y+y-z\| \leq \|x-y\| + \|y-z\|$$
$$\|x-z\| \leq \|x-y\| + \|y-z\|$$



3. Let A $\in$ $\mathrm{M}_{n*n}(\mathbb{R})$ be a matrix. AA^T is clearly symmetric but is it positive definite? AA^T is not necessarily positive definite, but it is positive semi-definite, meaning that

$$\langle x, AA^T, x\rangle \geq 0$$

for all vectors x. To see this, note that

$$\langle x, AA^T, x\rangle = \langle A^T x, A^T x\rangle = \left\|A^T x\right\|^2 \geq 0$$

A counter example to positive definiteness is provided, when n=2, by taking

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix};$$

then

$$AA^T = A$$

so if x $= (0,1)^T$,

$$\langle x, AA^T, x\rangle = 0.$$

Therefore cannot be strictly positive definite.

4. AA^T if is positively semi-definite $<->$ it is also right $A^T A$

It is true that $A^T A$ is symmetric. Let x be a non-zero column vector. Then we have: $x^T A^T Ax = (Ax)^T(Ax)$.

If we notice that Ax is also a non-zero column vector so $(Ax)^T(Ax)$ is the square of the inner product of Ax

Then: $(Ax)^T(Ax) = \|Ax\|^2 \geq 0$

5.

$$\|x\|_Q = \sqrt{x^2 Qx}$$

This is Frobenius norm which is the same as element-wize 2-norm. Also could be know as Hilbert-Schmidt norm, possibly infinite-dimesional Hilbert space. The Frobenius norm is an extension of the Euclidian norm to $K^{(nxn)}$ and comes from the Frobenious inner product on the space of all matrices. This norm is defined with formula below:

$$
\begin{aligned}
\|x\|_Q &= \sqrt{\sum_{i=0}^{m}(\sum_{j=1}^{n}|a_{ij}|^2)} \\
&= \sqrt{trace(A * A)} \\
&= \sqrt{\sum_{j=1}^{min(m,n)} \sigma_i^2(A)}
\end{aligned}
$$

Where $\delta_i(A)$ are singular values of A. Recall that trace function returns the sum of diagonal entries of square matrix.

4 Matrix Calculus

1.
$$f(x) = \frac{1}{2}\left\|Ax - b\right\|^2 \quad \text{for } x \in \mathbb{R}^n, b \in \mathbb{R}^m \text{ and } A \in \mathbb{R}^{m*n}$$

First step is to write equation in vector form as:

$$f(x) = \frac{1}{2}\left\|\begin{bmatrix} \sum_{j=0}^{n} a_{0j}x_j - b_0 \\ \vdots \\ \sum_{j=0}^{n} a_{mj}x_j - b_m \end{bmatrix}\right\|^2$$

Square root from norm is canceled by power of 2

$$f(x) = \frac{1}{2}\sum_{i=0}^{m}(\sum_{j=0}^{n} a_{ij}x_j - b_i)^2$$

next step is to derive values in side of the sum.

$$\nabla f(x) = \frac{\partial f(x)}{\partial x_k}$$
$$\frac{\partial f(x)}{\partial x_k} = \sum_{i=0}^{m}(\sum_{j=0}^{n} a_{ij}x_j - b_i)a_{ik}$$

When we transfer result to linear form we get:

$$\frac{\partial f(x)}{\partial x} = A^t(Ax - b)$$

2.
$$f(\alpha) = \frac{1}{2}\left\|A(x - \alpha y) - b\right\|^2$$

First step is to write equation in vector form as:

$$f(x) = \frac{1}{2}\left\|\begin{bmatrix} \sum_{j=0}^{n} a_{0j}(x_j + \alpha y_j) - b_0 \\ \vdots \\ \sum_{j=0}^{n} a_{mj}(x_j + \alpha y_j) - b_m \end{bmatrix}\right\|^2$$

Square root from norm is canceled by power of 2:

$$f(x) = \frac{1}{2}\sum_{i=0}^{m}(\sum_{j=0}^{n} a_{ij}(x_j + \alpha y_j) - b_i)^2$$

Now we calculate derivation inside of the sum:

$$\nabla f(\alpha) = \frac{\partial f(\alpha)}{\partial \alpha}$$
$$\frac{\partial f(\alpha)}{\partial \alpha} = \sum_{i=0}^{m}[(\sum_{j=0}^{n}(a_{ij}x_j + a_{ij}y_j\alpha) - b_i)(\sum_{j=0}^{n} a_{ij}y_j]$$

When we transfer result to linear form we get:

$$\frac{\partial f(\alpha)}{\partial \alpha} = y^t A^t (A(x + \alpha y) - b)$$

5 Student Task Selection Problem

1. We need to cast the student task selection problem into a linear program.
   We need to minimize the time which both students together will spend to finish the assignment. Our objective function is:

   $$Z = \sum_{i=1}^{15} \sum_{j=1}^{2} (x * t_{i,j})$$

   Where:

   $$\begin{aligned} Z &= \text{sum of times} \\ i &= \text{assignment number} \\ j &= \text{student number} \\ x &= 1 \text{ or } 0\text{, depending on the fact if student is working on i-th assignment or not} \\ t &= \text{time value for each assignment and student combination} \end{aligned}$$

   Both students have time budget, which is a maximum time they can spend on working on the assignment. Maximum time for the first student is 9, and for the second one is 6. Those values will be used to represent constraints, since they limit us. Writing that mathematically:

   $$\sum_{i=1}^{15} (x * t_{i,j}) \leq 9$$
   $$\sum_{i=1}^{15} (x * t_{i,j}) \leq 6$$

   In addition to that, it is important that both students do not work on the same assignment. We will write that constraint mathematically:

   $$x_{i,1} + x_{i_2} = 1 \quad \text{, where } i = 1, 2, 3, \ldots, 15$$

   This ensures that just one x value is 1, while the other is 0. Variable i is an assignment number, and 1 or 2 determines if it is for the first or the second student.

2. A linear program is in standard form if the following are all true:

   (a) Non-negativity constraints for all variables

   (b) All remaining constraints are expressed as equality constraints

   (c) The right hand side vector, b, is non-negative

With our linear program, the only thing we need to do is express the inequality constraints as equality constraints. To do that, we need to introduce the *slack variable* s. Slack variable measures the amount of "unused resource". It is used with inequalities where the unknown variable is less than given value. Since both of our inequalities are "less or equal" the given value, value of s can be 0 or greater.

Our newly written constraints are:

$$\sum_{i=1}^{15}(x * t_{i,1}) + s_1 = 9 \quad \text{, where } s_1 \geq 0$$
$$\sum_{i=1}^{15}(x * t_{i,2}) + s_2 = 6 \quad \text{, where } s_2 \geq 0$$

3. The python code for solving the linear program is in file *studentTaskSelection.py.*

   We used *scipy.optimize.linprog* to minimize the function Z. The method used is '*revised simplex*' because it is, according to the official documentation, more accurate than the default '*interior-point*'. It solves the given linear programming problem via a two-phase revised simplex algorithm.

   We needed to create variables which we can use as an input values for the *linprog* function, which can be seen in the code.

   The result of the function shows us the minimal time needed to complete all the assignment, as well as the distribution of the students.

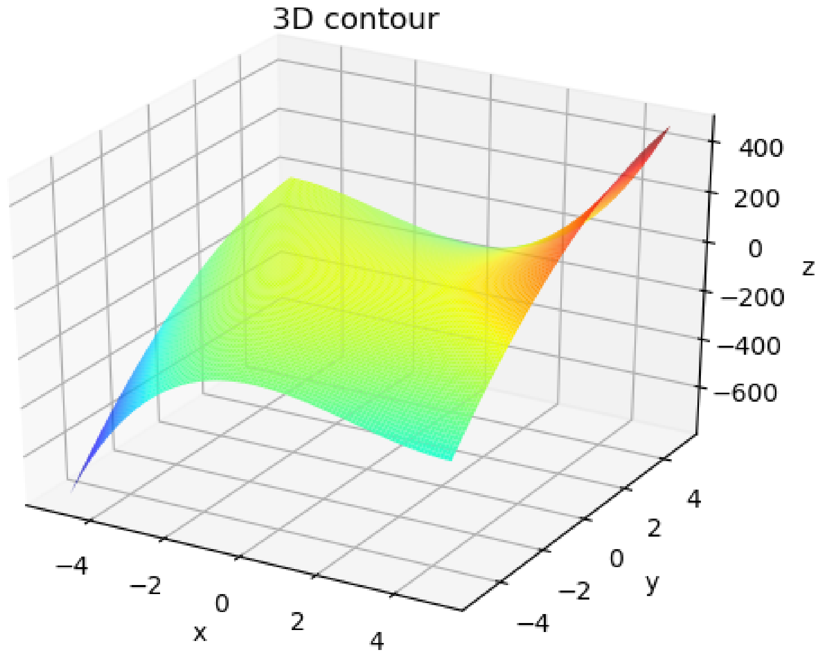
   The minimal time is 13.
   Student 1 will do tasks number: 1, 2, 3, 4, 9, 10, 12, 13 and 14, for which they will need 8.75 (hours, if time is shown in hours).
   Student 2 will do tasks number: 5, 6, 7, 8, 11, 15, for which they will need 4.25 hours.

4. All the steps are written under sub-tasks above.

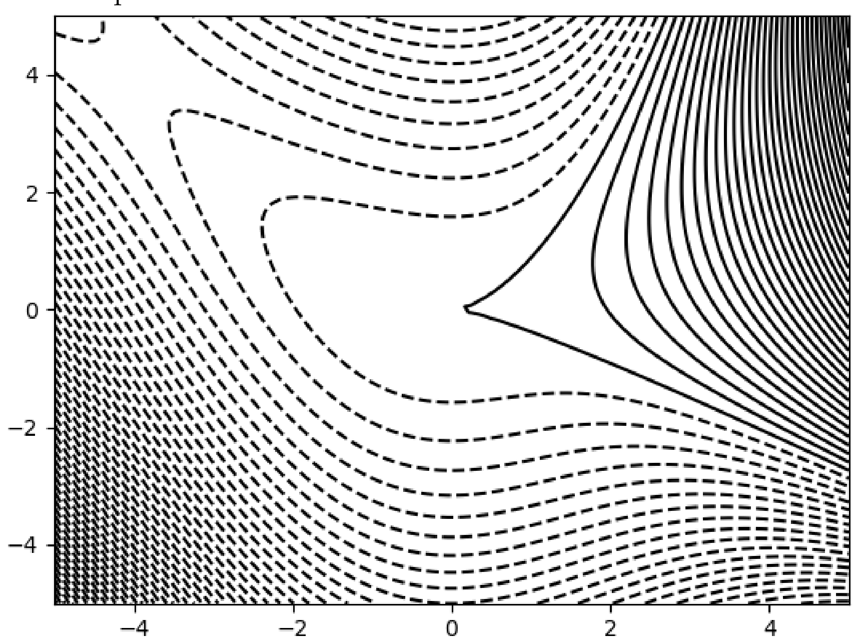
6 Addendum

1. Plots for functions in task 1:

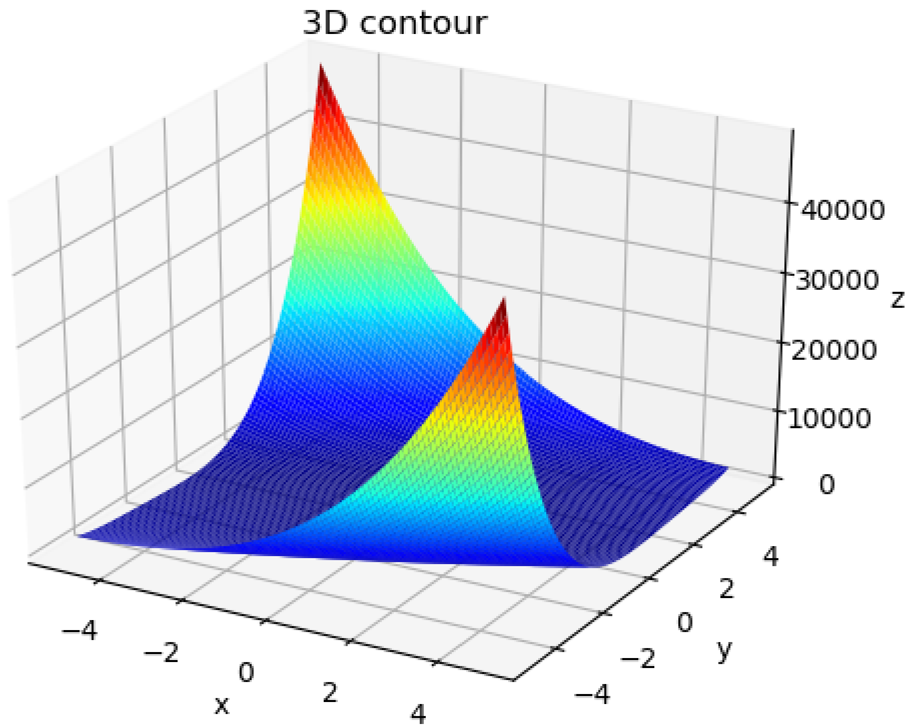
   (a) First function
   Contour plot



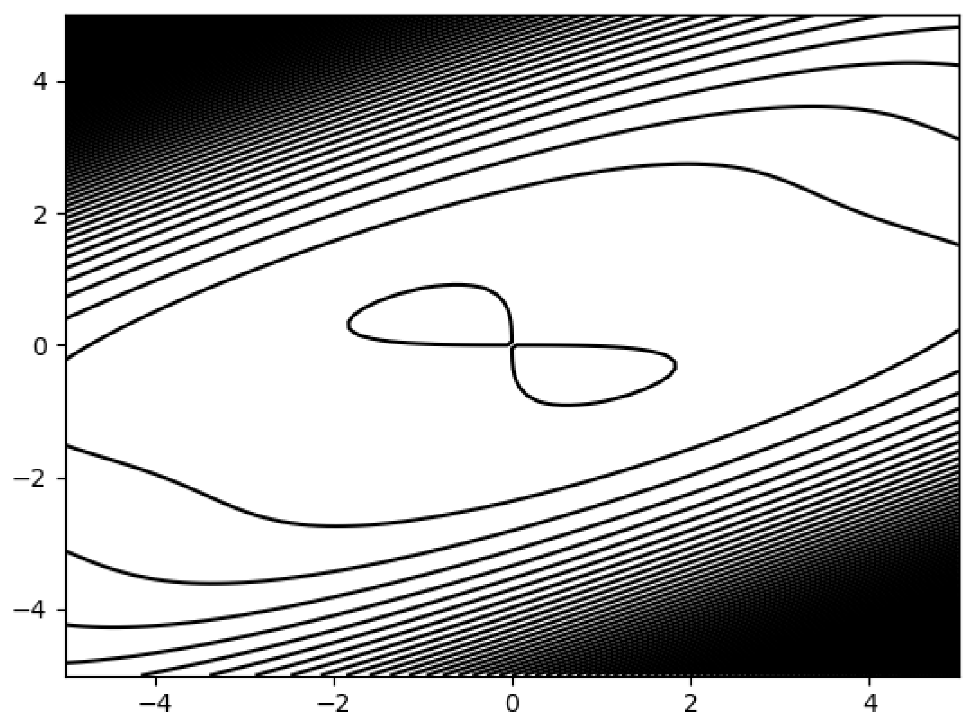
Surface plot



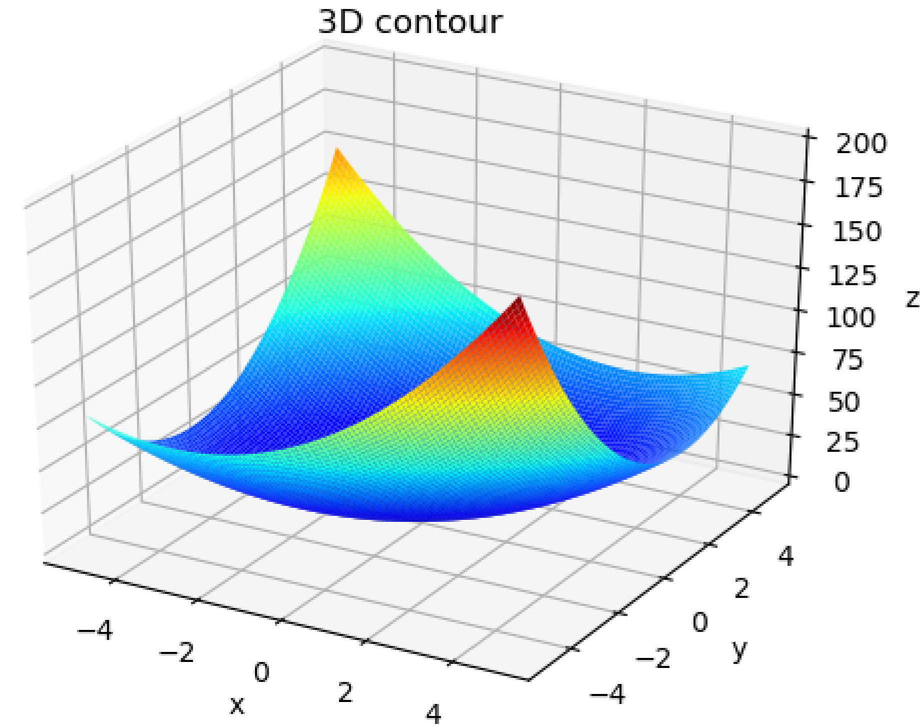
(b) Second function
Contour plot



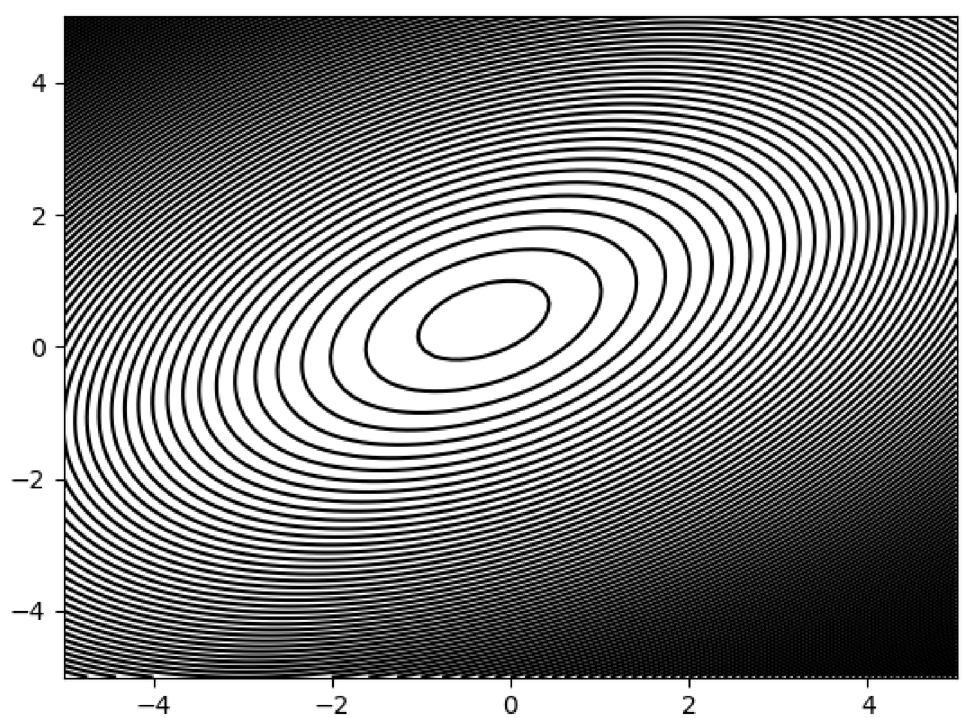
Surface plot



(c) Third function
Contour plot

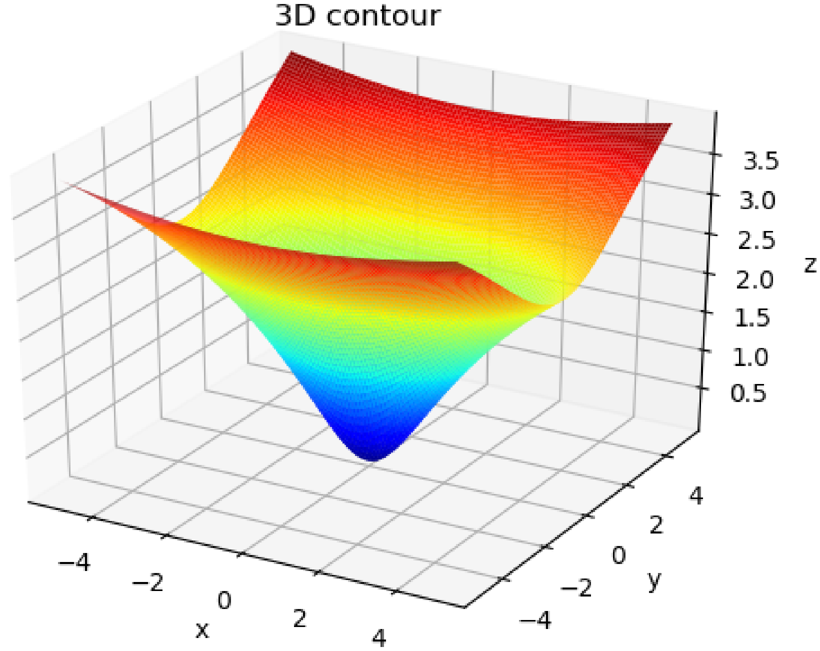


Surface plot

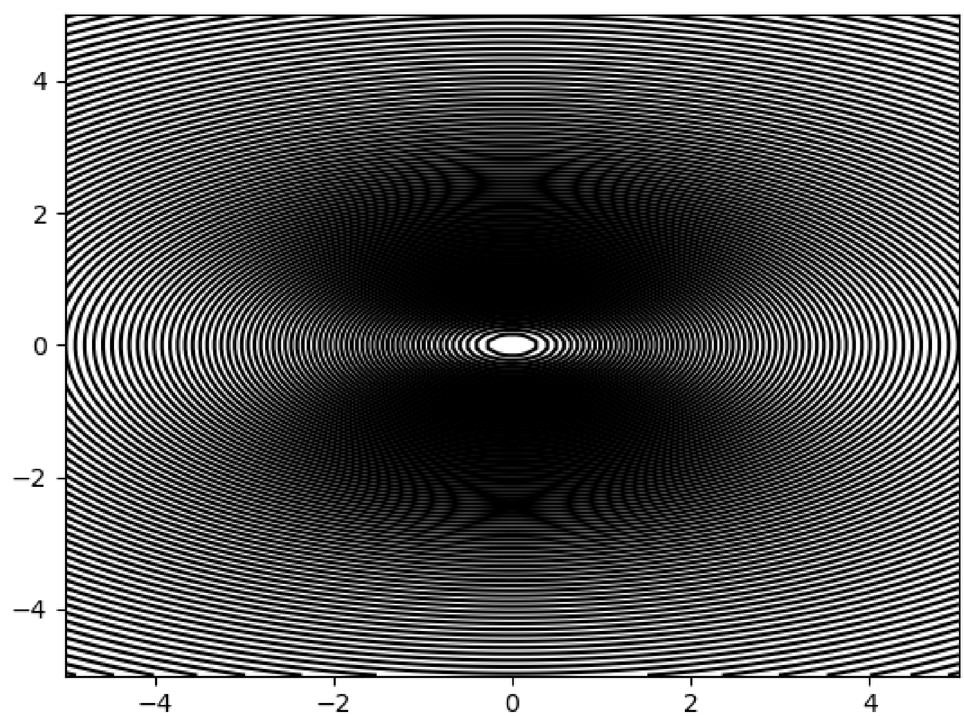


(d) Fourth function

Contour plot



Surface plot



2. Calculation for task 4:

4.1 $f(x) = \frac{1}{2}\|Ax - b\|^2 \qquad x \in R^n, \; b \in R^m, \; A \in R^{m\times n}$

$$f(x) = \frac{1}{2}\left\| \underbrace{\begin{bmatrix} a_{00} & \cdots & a_{0n} \\ \vdots & \ddots & \vdots \\ a_{m0} & \cdots & a_{mn} \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix}}_{n} - \underbrace{\begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}}_{m} \right\|^2 = \frac{1}{2}\left\| \underbrace{\begin{bmatrix} \sum_j^n a_{0j} x_j \\ \vdots \\ \sum_j^n a_{mj} x_j \end{bmatrix}}_{\in R^m} - \underbrace{\begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}}_{m} \right\|^2 =$$

$$= \frac{1}{2}\left\| \underbrace{\begin{bmatrix} \sum_j^n a_{0j} x_j - b_0 \\ \\ \sum_j^n a_{mj} x_j - b_m \end{bmatrix}}_{\text{2-VECTOR}} \right\|^2 \; * \qquad \nabla f(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_0} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \qquad \nabla f(x) \in R^n$$

$$f(x) = \frac{1}{2}\left\| \sum_i^n z_i \right\|^2 = \frac{1}{2}\sqrt{\left(\sum_i^n z_i^2\right)}^{\cancel{2}} = \frac{1}{2}\sum_i^n z_i^2 \xrightarrow[\nabla f(x) = \frac{\partial f(x)}{\partial x}]{} \cancel{\frac{1}{2}} \sum_i^n \cancel{2} z_i$$

$*$

$$\nabla f = \frac{1}{2}\sqrt{\left[\sum_i^m \left(\sum_j^n a_{ij} x_j - b_i\right)^2\right]}^{2} = \cancel{\frac{1}{2}} \sum_j^m \left[\cancel{2} \sum_j^n (a_{ij} x_j - b_i) \cdot (a_{ik} x_k - b_i)'\right]$$

$x_j = x_k$

$$\nabla f = \frac{\partial f(x)}{\partial x_k} = \sum_i^m \left[\underbrace{\sum_j^n (a_{ij} x_j - b_i)}_{n} \underbrace{a_{ik}}_{m\times n}\right] = \sum_i^m \left[\sum_j^n \underbrace{a_{ik}}_{A^T} \underbrace{(a_{ij} x_j - b_i)}_{(Ax+b)}\right]$$

$$= A^T (Ax + b)$$

4.2 $f(\alpha) = \frac{1}{2} \| A(x+\alpha y) - b \|^2 \qquad \alpha \in \mathbb{R},\ x, y \in \mathbb{R}^n,\ b \in \mathbb{R}^m,\ A \in \mathbb{R}^{m \times n}$

$$f(\alpha) = \frac{1}{2} \left\| \begin{bmatrix} a_{00} & \cdots & a_{0n} \\ \vdots & & \vdots \\ a_{m0} & \cdots & a_{mn} \end{bmatrix} \left(\begin{bmatrix} x_0 \\ \vdots \\ x_n \end{bmatrix} + \alpha \begin{bmatrix} y_0 \\ \vdots \\ y_n \end{bmatrix} \right) - \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix} \right\|^2 =$$

$$= \frac{1}{2} \left\| \underbrace{\begin{bmatrix} a_{00} & \cdots & a_{0n} \\ \vdots & & \vdots \\ a_{m0} & \cdots & a_{mn} \end{bmatrix}}_{m \times n} \left(\underbrace{\begin{bmatrix} x_0 + \alpha y_0 \\ \vdots \\ x_n + \alpha y_n \end{bmatrix}}_{n} \right) - \underbrace{\begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix}}_{n} \right\|^2 = \frac{1}{2} \left\| \begin{bmatrix} \sum_j^n a_{0j}(x_j + \alpha y_j) \\ \vdots \\ \sum_j^n a_{mj}(x_j + \alpha y_j) \end{bmatrix} - \begin{bmatrix} b_0 \\ \vdots \\ b_n \end{bmatrix} \right\|^2$$

$$= \frac{1}{2} \left\| \begin{bmatrix} \left(\sum_j^n a_{0j}(x_j + \alpha y_j)\right) - b_0 \\ \vdots \\ \left(\sum_j^n a_{mj}(x_j + \alpha y_j)\right) - b_n \end{bmatrix} \right\|^2 \qquad \nabla f(x) = \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_0} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$\nabla f = \frac{1}{2} \sqrt{\left[\sum_i^m \left(\sum_j^n a_{ij}(x_j + \alpha y_j) - b_i \right) \right]^2}^{\,2} \underset{x_j = x_k}{=} \frac{1}{2} \sum_i^m \left[\sum_j^n \left(a_{ij}(x_j + \alpha y_j) - b_i \right) \right]^2$$

$$\underset{\frac{\partial f(\alpha)}{\partial \alpha}}{\Longrightarrow} \cancel{\tfrac{1}{2}}\, 2 \sum_i^m \sum_j^n \left[(a_{ij} x_j + a_{ij} \alpha y_i) - b_i \right] \cdot \left[(a_{ij} x_j + \alpha a_{ij} y_i) - b_i \right]'$$

$$= \sum_i^m \sum_j^n \left[(a_{ij} x_j + a_{ij} \alpha y_i) - b_i \right] \left[a_{ij} y_i \right]$$

$$= \sum_i^m \sum_j^n \left[a_{ij} y_i \right] \left[a_{ij} (x_j + \alpha y_i) - b_i \right] \Longrightarrow y^T A^T \left(A(x + \alpha y) - b \right)$$