Optimization

Optimization

- Optimization means finding that value of **x** which maximizes or minimizes a given function *f (***x***)*.
- Closely related problem is that of solving a nonlinear equation,**g***(***x***)* = **0** for **x**, where **g** is a possibly multivariate function.
- An example of a multivariate function:

$$f(x_1, x_2, x_3) = x_1^{\ 4} + x_1 x_2 x_3 + 2x_2 x_3^{\ 2}$$

Multidimensional Gradient Methods

- Use information from the derivatives of the optimization function to guide the search
- Finds solutions quicker compared with direct search methods
- A good initial estimate of the solution is required
- The objective function needs to be differentiable

Taylor Expansion

$$f(x+h) = f(x) + hf^{'}(x) + \frac{h^2}{2!} f^{''}(x) + \frac{h^3}{3!} f^{'''}(x) + \frac{h^4}{4!} f^{''''}(x)\ldots$$

$$F(x+h) = F(x) + h^T g + \frac{1}{2} h^T Hh + 0(\|h^3\|)$$

$$H_s = F''(x_s) \rightarrow$$

if the second derivatives of *f* are all continuous in a neighborhood *D*, then the Hessian of *f* is a symmetric matrix throughout *D*.

In linear algebra, a **symmetric matrix** is a **square matrix** that is **equal to its transpose**.

Formally, matrix *A* is symmetric if $A = A^T$ and $a_{ij} = a_{ji}$

Gradients

- The *gradient* is a vector operator denoted by ∇ (referred to as "del")
- When applied to a function , it represents the functions directional derivatives
- The gradient is the special case where the direction of the gradient is the direction of most or the *steepest ascent/descent*
- The gradient is calculated by

$$\nabla f = \frac{\partial f}{\partial x}\mathrm{i} + \frac{\partial f}{\partial y}\mathrm{j}$$

Gradients-Example

Calculate the gradient to determine the direction of the steepest slope at point (2, 1) for the function $f(x, y) = x^2 y^2$

Solution: To calculate the gradient we would need to calculate which are used to determine the gradient at point (2,1) as

$$\frac{\partial f}{\partial x} = 2xy^2 = 2(2)(1)^2 = 4 \qquad \frac{\partial f}{\partial y} = 2x^2 y = 2(2)^2(1) = 8$$

$$\nabla f = 4\mathrm{i} + 8\mathrm{j}$$

Gradient Vector

In the three-dimensional Cartesian coordinate system, the gradient is calculated by

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

where **i, j, k are the standard unit vectors**. For example, the gradient of the function

$$f(x, y, z) = 2x + 3y^2 - \sin(z)$$

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k} = 2\mathbf{i} + 6y\mathbf{j} - \cos(z)\mathbf{k}.$$

Gradient Vector

In some applications it is customary to **represent the gradient as a row vector or column vector of its components** in a rectangular coordinate system:

The gradient of a function of *n* variables $f(x_1, x_2, \dots, x_n)$ is defined as follows:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots\dots, \frac{\partial f}{\partial x_n}\right)^T = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \dots\dots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

Gradient Example

For the function $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{16}\boldsymbol{x}_{\mathbf{1}} + \mathbf{12}\boldsymbol{x}_{\mathbf{2}} + \boldsymbol{x}_{\mathbf{1}}^{\mathbf{2}} + \boldsymbol{x}_{\mathbf{2}}^{\mathbf{2}}$ the gradient :

$$\boldsymbol{\nabla f}(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{\mathbf{1}}} \\ \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{\mathbf{2}}} \end{pmatrix} = \begin{pmatrix} \mathbf{16} + \mathbf{2}\boldsymbol{x}_{\mathbf{1}} \\ \mathbf{12} + \mathbf{2}\boldsymbol{x}_{\mathbf{2}} \end{pmatrix}$$

For the function $f(x, y, z) = 2x + 3y^2 - \sin(z)$ the gradient :

$$\nabla f(x) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)^T = (2, 6y, -\cos(z))^T$$

Jacobian Matrix

In vector calculus, the **Jacobian matrix** is the matrix of all first-order partial derivatives of a vector-valued function.

$$F: \mathbb{R}^n \to \mathbb{R}^m$$

$$J = \begin{bmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1} & \cdots & \dfrac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

In the case ***m=n*** the Jacobian matrix is a square matrix

$$F(x,y) = \begin{bmatrix} x^2 y \\ 5x + \sin(y) \end{bmatrix}. \qquad F_1(x,y) = x^2 y \qquad F_2(x,y) = 5x + \sin(y)$$

and the Jacobian matrix of *F* is

$$J_F(x,y) = \begin{bmatrix} \dfrac{\partial F_1}{\partial x} & \dfrac{\partial F_1}{\partial y} \\ \dfrac{\partial F_2}{\partial x} & \dfrac{\partial F_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 2xy & x^2 \\ 5 & \cos(y) \end{bmatrix}$$

and the Jacobian determinant is

$$\det(J_F(x,y)) = 2xy\cos(y) - 5x^2.$$

Hessian

- The *Hessian* matrix or just the *Hessian* is the Jacobian matrix of second-order partial derivatives of a function.
- The determinant of the Hessian matrix is also referred to as the Hessian.
- For a two dimensional function the Hessian matrix is simply

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

Hessians

The Hessian matrix of a function of n variables $f(x_1, x_2, \dots., x_n)$ is as follows :

$$H = \nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right] \qquad i, j = 1,2,\dots,n$$

$$H = f''(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] = \nabla^2 f(x) = \begin{bmatrix} \partial^2 f / \partial x_1{}^2 & \partial^2 f / \partial x_1 \partial x_2 \dots. & \partial^2 f / \partial x_1 \partial x_n \\ \partial^2 f / \partial x_2 \partial x_1 & \partial^2 f / \partial x_2{}^2 \dots. & \partial^2 f / \partial x_2 \partial x_n \\ \dots\dots\dots\dots\dots & \dots\dots\dots\dots\dots & \dots\dots\dots\dots \\ \partial^2 f / \partial x_n \partial x_1 & \dots\dots\dots\dots\dots & \partial^2 f / \partial x_n{}^2 \end{bmatrix}$$

Hessians-Example

Calculate the Hessian matrix for the function

$$f(x, y) = x^2 y^2$$

To calculate the Hessian matrix; the partial derivatives must be evaluated as

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

resulting in the Hessian matrix

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2y^2 & 4xy \\ 4xy & 2x^2 \end{bmatrix}$$

First Derivative Test For Local Extreme Values

Let x^* be an interior point of a domain *D* in R^n and assume that f is twice continously differentiable on *D*. *It is necessary for a local minimum or a maximum of f at x^* that:*

$$\nabla f(x^*) = 0$$

This implies that

$$\frac{\partial f(x^*)}{\partial x_1} = 0$$

$$\frac{\partial f(x^*)}{\partial x_2} = 0$$

$$\vdots$$

$$\frac{\partial f(x^*)}{\partial x_n} = 0$$

First Derivative Test For Local Extreme Values

First derivative test is inconclusive to find a local extramum point.

Critical points and saddle points.

Critical point. An interior point of the domain of a function $f(x_1, x_2, \ldots, x_n)$ where all first partial derivatives are zero or where one or more of the first partials does not exist is a **critical point** of f.

Saddle point. A critical point that is not a local extremum is called a saddle point. We can say that a differentiable function $f(x_1, x_2, \ldots, x_n)$ has a saddle point at a critical point (x^*) if we can partition the vector x^* into two subvectors (x^{1*}, x^{2*}) where $x^{1*} \in X^1 \subseteq R^q$ and $x^{2*} \in X^2 \subseteq R^p$ $(n = p + q)$ with the following property

$$f(x^1, x^{2*}) \leq f(x^{1*}, x^{2*}) \leq f(x^{1*}, x^2)$$

Second Derivative Test For Local Extreme Values

The determinant of the Hessian matrix denoted by $|H|$ can have three cases:

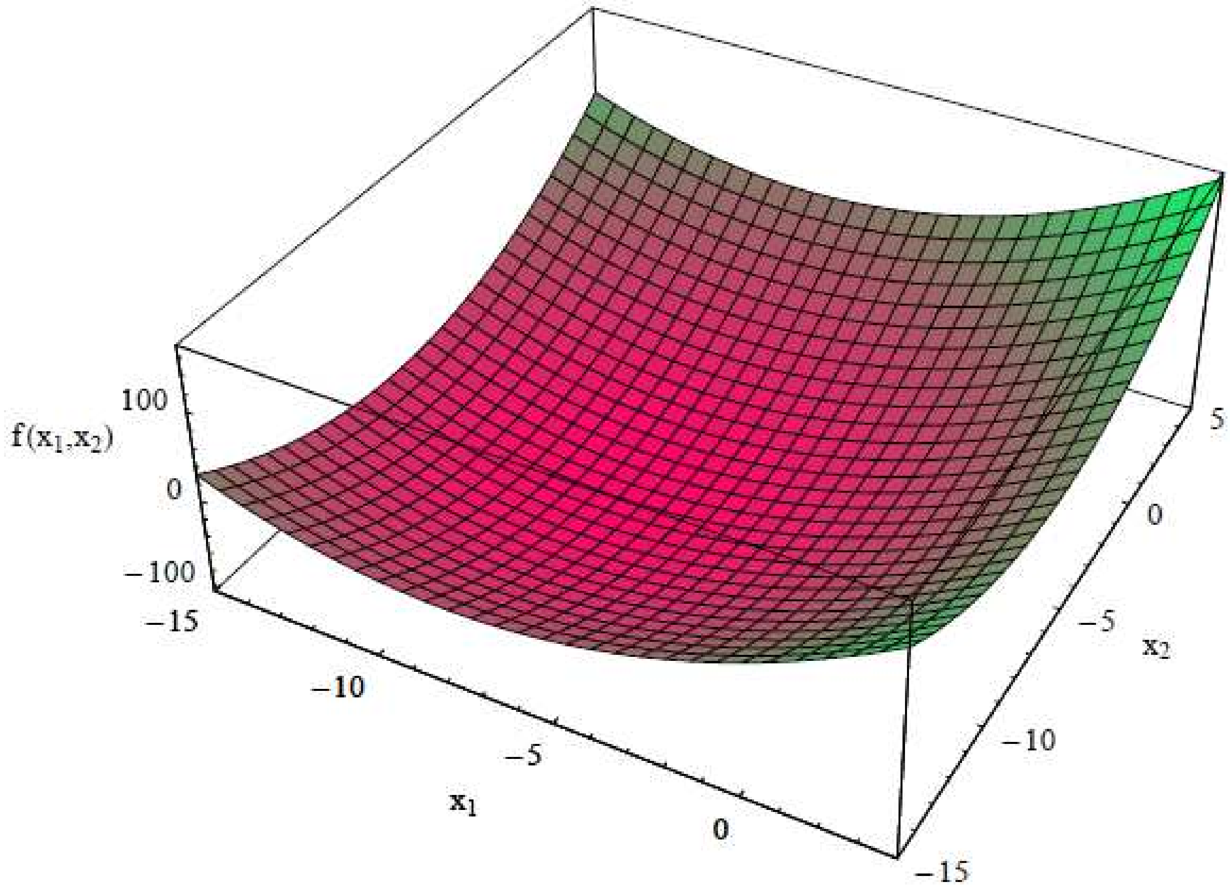
1. If $|H| > 0$ and $\partial^2 f / \partial x^2 > 0$ then $f(x, y)$ has a local minimum.
2. If $|H| > 0$ and $\partial^2 f / \partial x^2 < 0$ then $f(x, y)$ has a local maximum.
3. If $|H| < 0$ then $f(x, y)$ has a saddle point.

Euclidean Norm

On an n-dimensional Euclidean space $\mathbf{R}^n$, the intuitive notion of length of the vector $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is captured by the formula:

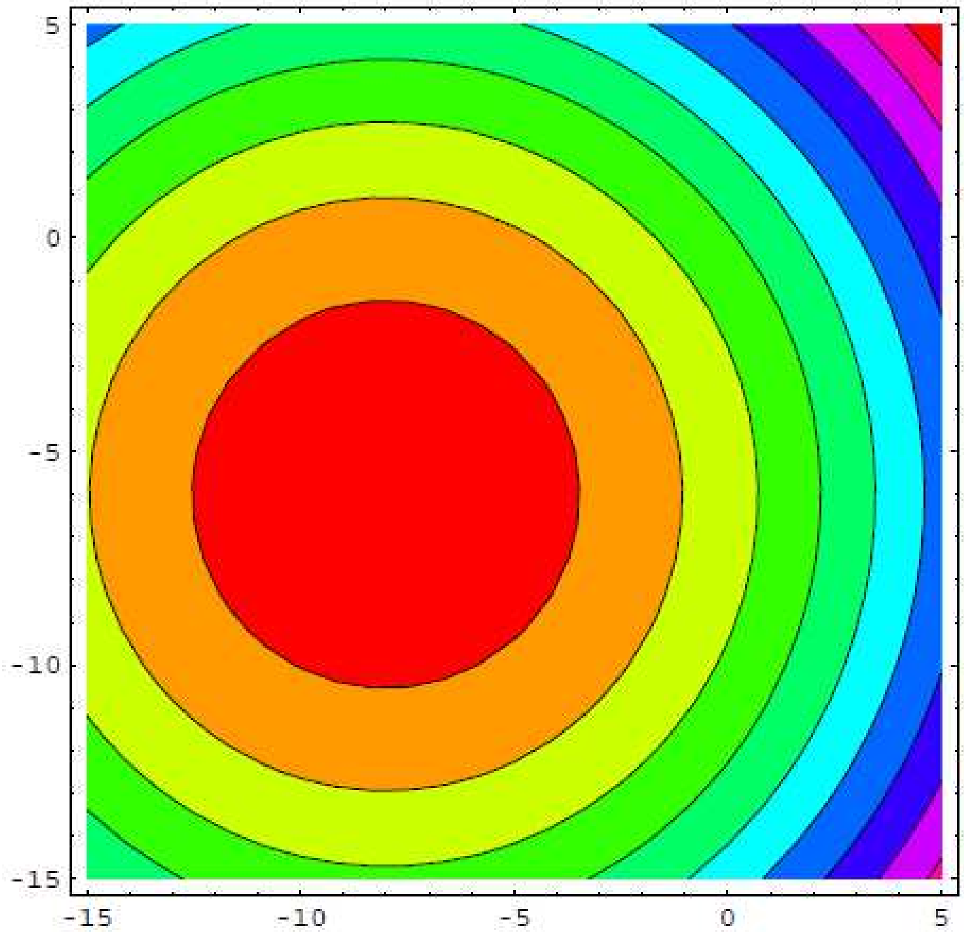
$$\|\boldsymbol{x}\| := \sqrt{x_1^2 + \cdots + x_n^2}.$$

Graph of the function $f(x_1, x_2) = 16x_1 + 12x_2 + x_1^2 + x_2^2$

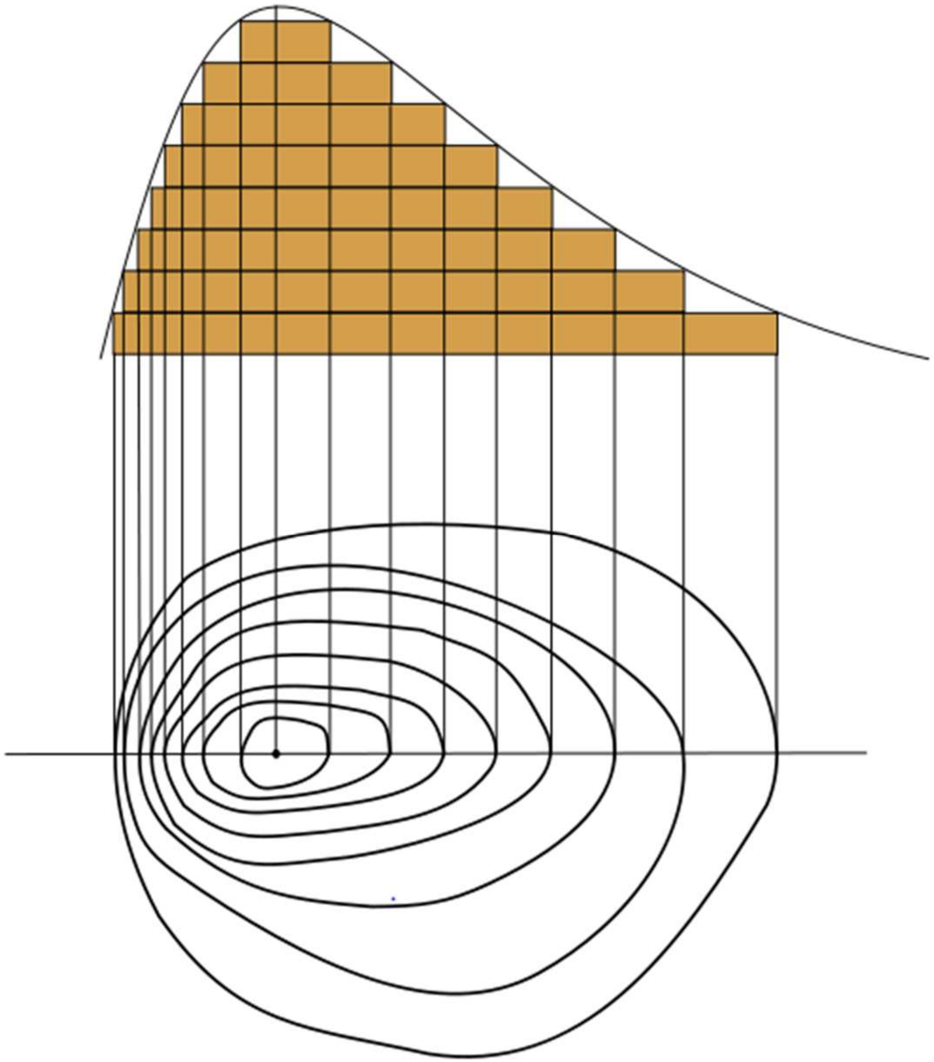


Contour or Level Curves

Level curves of the function $16x_1 + 12x_2 + x_1^2 + x_2^2$

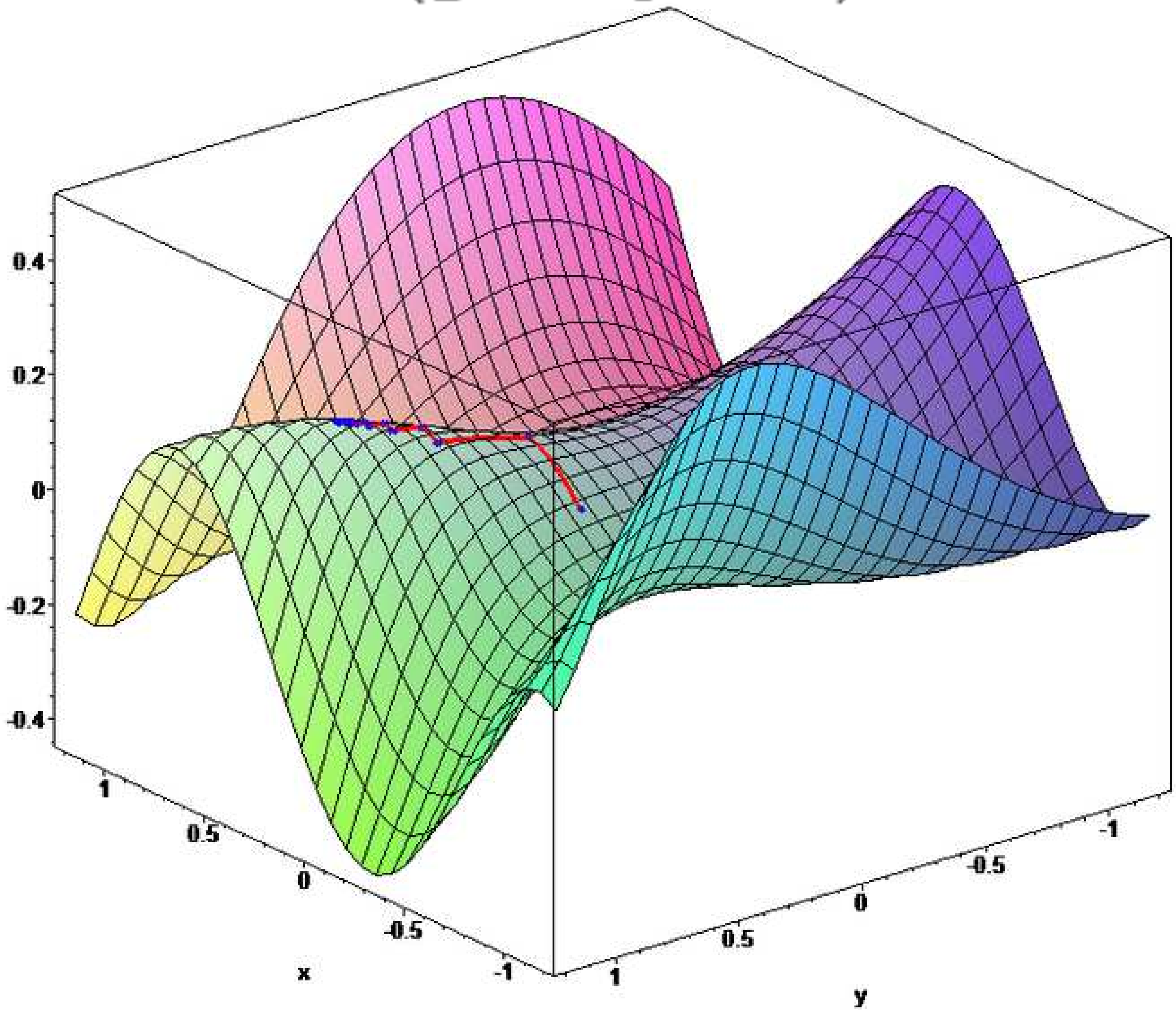


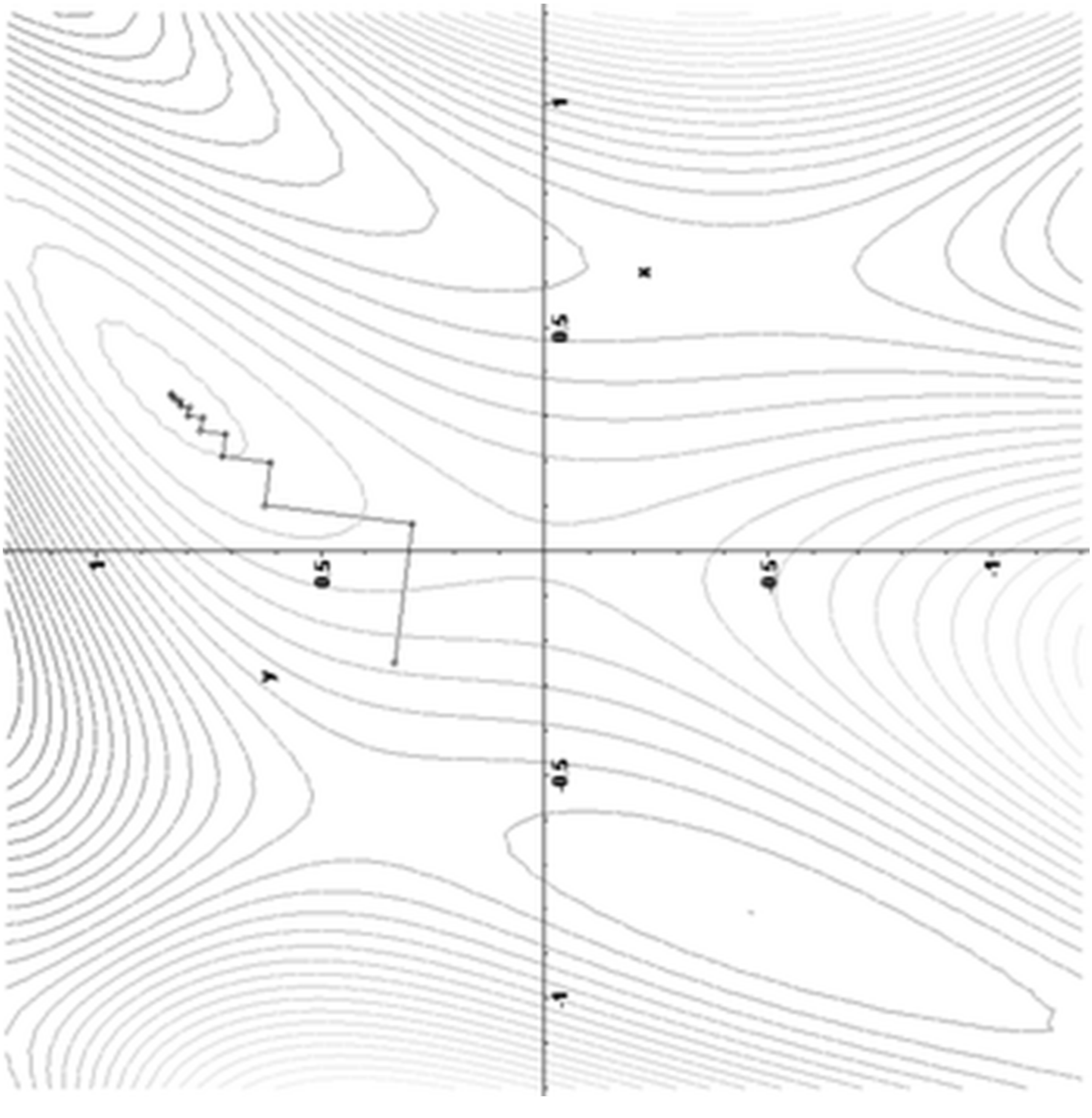
A contour line (often just called a "contour") joins points of equal elevation (height) above a given level,



$$F(x,y) = \sin\left(\frac{1}{2}x^2 - \frac{1}{4}y^2 + 3\right)\cos(2x + 1 - e^y)$$

F(x,y)



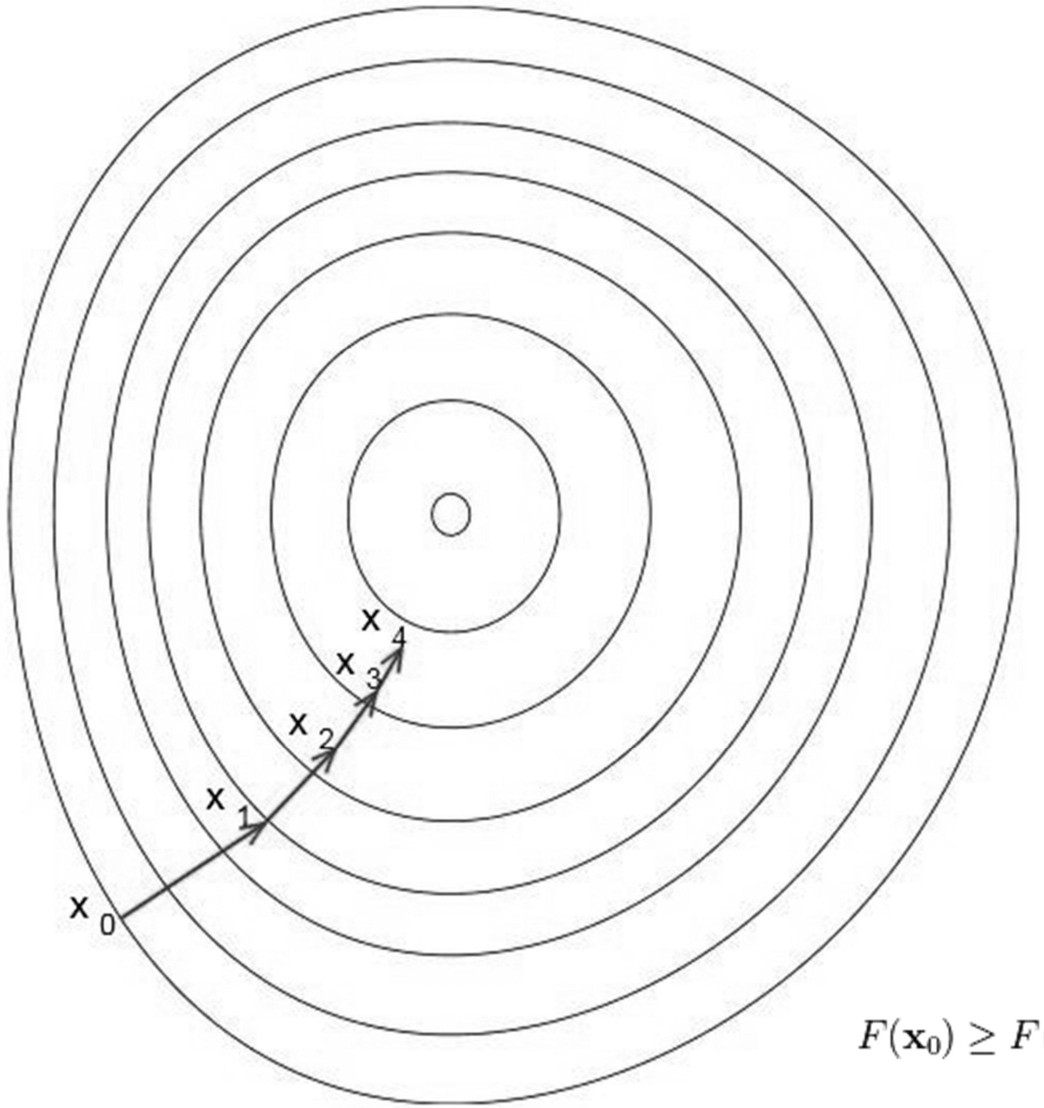


x
y
1
0.5
-0.5
-1

Gradient Descent Method

The gradient $\frac{\partial f(x)}{\partial x}$ at location x points toward a direction where the function increases. The negative $-\frac{\partial f(x)}{\partial x}$ is usually called *steepest descent direction*

Illustration of steepest descent



$$F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \cdots ,$$

Steepest Descent

Step 1: Set x^0, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, M - the maximum number of iterations.
Find the gradient $\nabla f(x)$ of the objective function.

Step 2: Set $k=0$.

Step 3: Compute $\nabla f(x^k)$.

Step 4: Check the stopping criteria $||\nabla f(x^k)|| < \varepsilon_1$:
a) if the condition is satisfied, set $x^* = x^k$ and finish the search process;
b) if the condition is not satisfied, go to step 5.

Step 5: Check the condition $k \geq M$:
a) if it is satisfied, finish the search process and set $x^* = x^k$;
b) if it is not satisfied, go to step 6.

Steepest Descent

Step 6: Find step length t_k minimizing the function $\varphi(t_k)=f(x^k-t_k\cdot\nabla f(x^k))$.

Step 7: Compute $x^{k+1}=x^k-t_k\cdot\nabla f(x^k)$.

Step 8: Check the finishing conditions:
$||x^{k+1}-x^k|| < \varepsilon_2$
$|f(x^{k+1})-f(x^k)|<\varepsilon_2$

a) if both conditions are satisfied with numbers k and k-1, finish the search process and set $x^*=x^{k+1}$;

b) if both conditions are not satisfied, set $k=k+1$ and go to step 3.

Steepest Descent Example (on board)