Unconstrained Optimization ND

What is the optimal solution? (ND)

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$$

(First-order) Necessary condition

1D: $f'(x) = 0$

(Second-order) Sufficient condition

1D: $f''(x) > 0$

Taking derivatives…

From linear algebra:

A symmetric $n \times n$ matrix $\boldsymbol{H}$ is **positive definite** if $\boldsymbol{y}^T \boldsymbol{H}\ \boldsymbol{y} > \boldsymbol{0}$ for any $\boldsymbol{y} \neq \boldsymbol{0}$

A symmetric $n \times n$ matrix $\boldsymbol{H}$ is **positive semi-definite** if $\boldsymbol{y}^T \boldsymbol{H}\ \boldsymbol{y} \geq \boldsymbol{0}$ for any $\boldsymbol{y} \neq \boldsymbol{0}$

A symmetric $n \times n$ matrix $\boldsymbol{H}$ is **negative definite** if $\boldsymbol{y}^T \boldsymbol{H}\ \boldsymbol{y} < \boldsymbol{0}$ for any $\boldsymbol{y} \neq \boldsymbol{0}$

A symmetric $n \times n$ matrix $\boldsymbol{H}$ is **negative semi-definite** if $\boldsymbol{y}^T \boldsymbol{H}\ \boldsymbol{y} \leq \boldsymbol{0}$ for any $\boldsymbol{y} \neq \boldsymbol{0}$

A symmetric $n \times n$ matrix $\boldsymbol{H}$ that is not negative semi-definite and not positive semi-definite is called **indefinite**

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$$

First order necessary condition: $\boldsymbol{\nabla} f(\boldsymbol{x}) = \mathbf{0}$

Second order sufficient condition: $\boldsymbol{H}(\boldsymbol{x})$ **is positive definite**

How can we find out if the Hessian is positive definite?

Types of optimization problems

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$$

f: nonlinear, continuous and smooth

Gradient-free methods

Evaluate $f(\boldsymbol{x})$

Gradient (first-derivative) methods

Evaluate $f(\boldsymbol{x}), \boldsymbol{\nabla} f(\boldsymbol{x})$

Second-derivative methods

Evaluate $f(\boldsymbol{x}), \boldsymbol{\nabla} f(\boldsymbol{x}), \boldsymbol{\nabla}^2 f(\boldsymbol{x})$

Example (ND)

Consider the function $f(x_1, x_2) = 2x_1^3 + 4x_2^2 + 2x_2 - 24x_1$

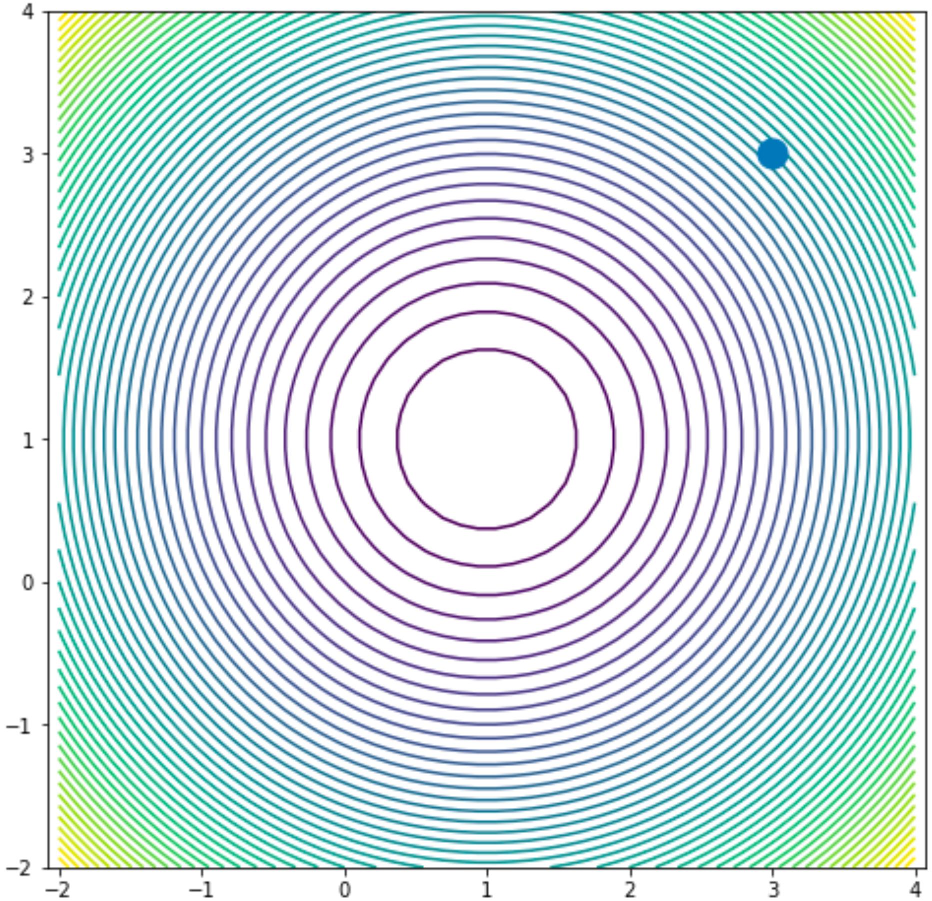
Find the stationary point and check the sufficient condition

Optimization in ND: Steepest Descent Method

Given a function $f(\boldsymbol{x}): \mathcal{R}^n \rightarrow \mathcal{R}$ at a point $\boldsymbol{x}$, the function will decrease its value in the direction of steepest descent: $-\boldsymbol{\nabla} f(\boldsymbol{x})$

What is the steepest descent direction?

$$f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$$



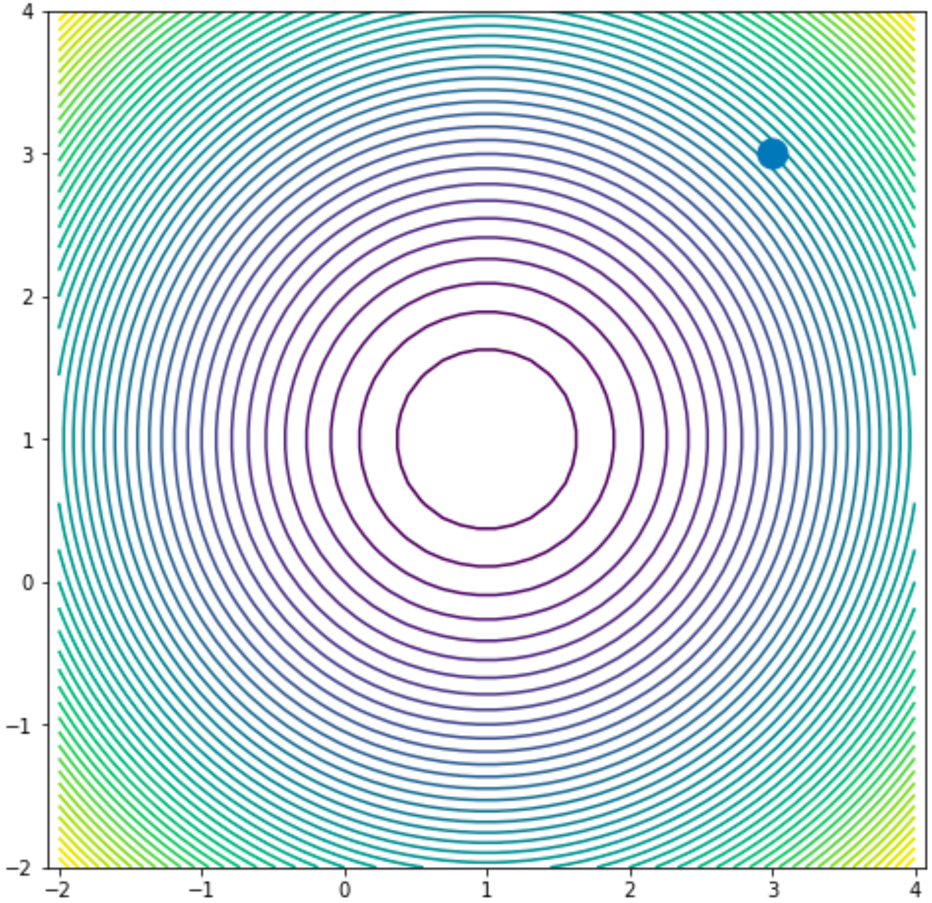
Steepest Descent Method

Start with initial guess:

$$\boldsymbol{x}_0 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Check the update:

$$f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$$



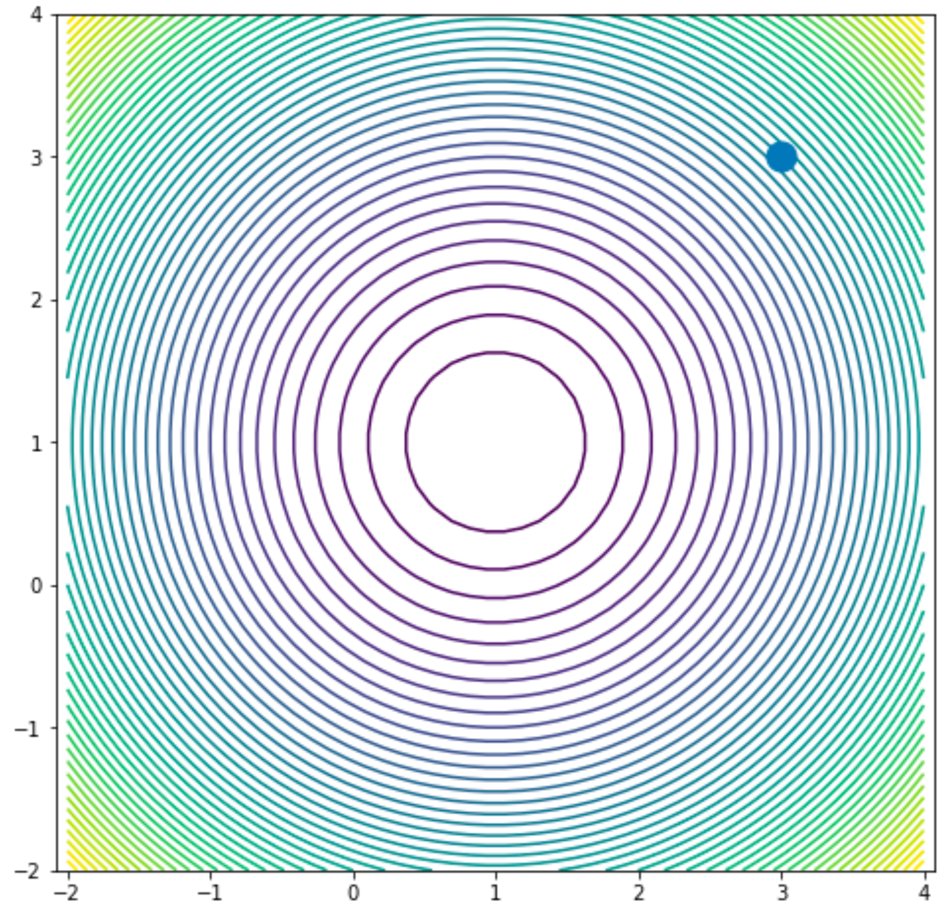
Steepest Descent Method

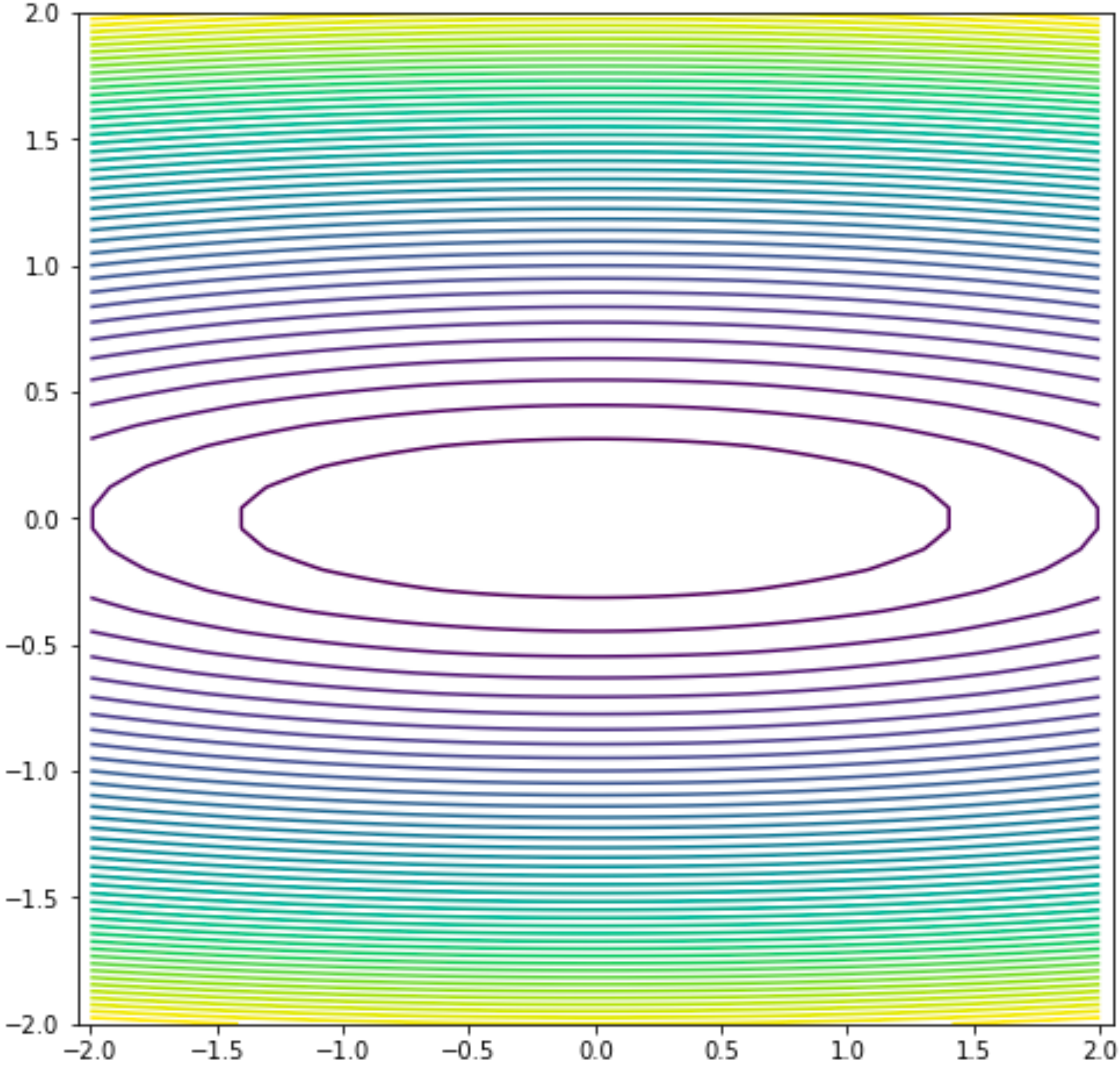
Update the variable with:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{\nabla} f(\boldsymbol{x}_k)$$

How far along the gradient should we go? What is the "best size" for α_k?

$$f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2$$




2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

Steepest Descent Method

Algorithm:

Initial guess: $\boldsymbol{x}_0$

Evaluate: $\boldsymbol{s}_k = -\boldsymbol{\nabla} f(\boldsymbol{x}_k)$

Perform a line search to obtain α_k (for example, Golden Section Search)

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} f(\boldsymbol{x}_k + \alpha\, \boldsymbol{s}_k)$$

Update: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k\, \boldsymbol{s}_k$

Line Search

Example

Consider minimizing the function

$$f(x_1, x_2) = 10(x_1)^3 - (x_2)^2 + x_1 - 1$$

Given the initial guess

$$x_1 = 2, x_2 = 2$$

what is the direction of the first step of gradient descent?

Newton's Method

Using Taylor Expansion, we build the approximation:

Newton's Method

Algorithm:

Initial guess: $\boldsymbol{x}_0$

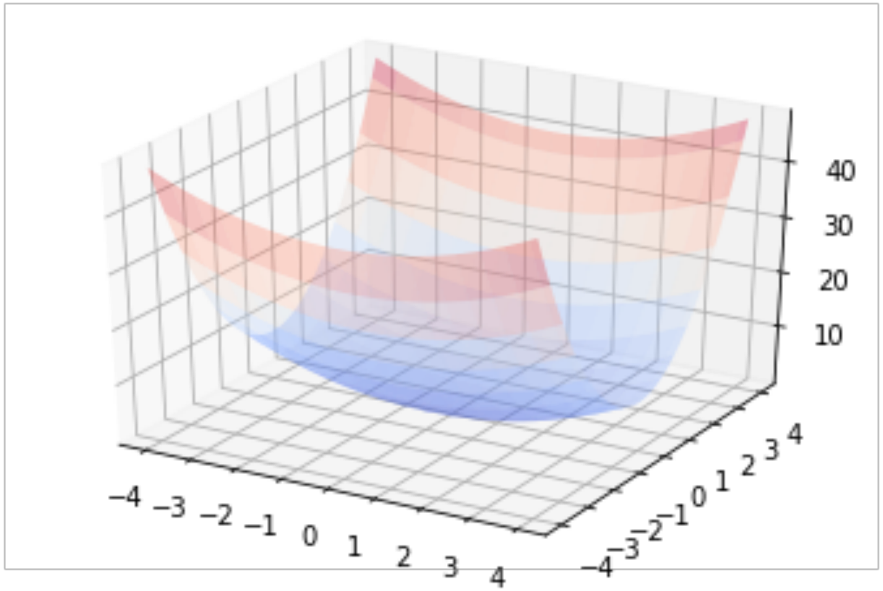
Solve: $\boldsymbol{H}_{\boldsymbol{f}}(\boldsymbol{x}_k)\,\boldsymbol{s}_k = -\boldsymbol{\nabla} f(\boldsymbol{x}_k)$

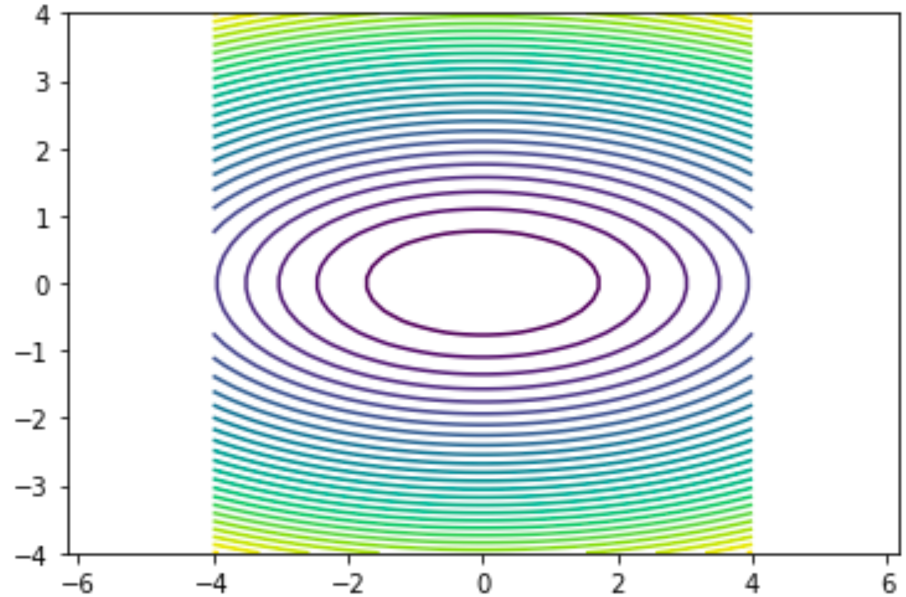
Update: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{s}_k$

Note that the Hessian is related to the curvature and therefore contains the information about how large the step should be.

Try this out!

$$f(x, y) = 0.5x^2 + 2.5y^2$$





When using the Newton's Method to find the minimizer of this function, estimate the number of iterations it would take for convergence?

A) 1 B) 2-5 C) 5-10 D) More than 10 E) Depends on the initial guess

Newton's Method Summary

Algorithm:

Initial guess: $\boldsymbol{x}_0$

Solve: $\boldsymbol{H}_{\boldsymbol{f}}(\boldsymbol{x}_k)\,\boldsymbol{s}_k = -\boldsymbol{\nabla} f(\boldsymbol{x}_k)$

Update: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{s}_k$

About the method…

- Typical quadratic convergence ☺
- Need second derivatives ☹
- Local convergence (start guess close to solution)
- Works poorly when Hessian is nearly indefinite
- Cost per iteration: $O(n^3)$