Homework Project

Victor DUC

December 3, 2022

The aim of this project is to write a program that implements the following method.

Algorithm 98 – Steepest descent algorithm (gradient descent)

$(S.0)$ Choose $x^0 \in U$ and let $k = 0$.

$(S.1)$ If $\nabla f(x^k) = 0$: STOP.

$(S.2)$ Determine an efficient stepsize $t_k > 0$ and let $x^{k+1} := x^k - t_k \nabla f(x^k)$.

$(S.3)$ Let $k := k + 1$ and go to $(S.1)$.

We will use is to minimize Rosenbrock's function and Easom's functions with starting point $x^0 := (0,0)^\top$ and then $\tilde{x}^0 := (\pi + 1, \pi - 1)^\top$ (for each function). We will first use a constant stepsize, then Armijo stepsize.

Definition 93 (Armijo-stepsize)
Let $x \in U$ and $d \in \mathbb{R}^n$ such that $\nabla f(x)^\top d < 0$.

The **Armijo-stepsize** $t_A \in \mathbb{R}$ is a solution of the system in t

$$\begin{cases} f(x+td) \leqslant f(x) + \delta t \nabla f(x)^\top d \\ t \geqslant -\sigma \frac{\nabla f(x)^\top d}{\|d\|^2} \end{cases}$$

for some constants $\delta \in]0,1[$ and $\sigma > 0$ independant of x and d.

Contents

1 Rosenbrock's function

Rosenbrock's function is differentiable.

$$
\begin{array}{rccl}
f: & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\
 & (x_1, x_2) & \longmapsto & (1-x_1)^2 + 100\left(x_2 - x_1^2\right)^2
\end{array}
$$

The gradient of f is given by

$$\forall (x_1, x_2) \in \mathbb{R}^2 \quad \nabla f(x_1, x_2) = \begin{pmatrix} 400x_1(x_1^2 - x_2) + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$

The point $(1,1)^\top$ is the only stationary point of f.

$$\{(x_1, x_2) \in \mathbb{R}^2 \mid \nabla f(x_1, x_2) = (0,0)^\top\} = \{(1,1)\}$$

Rosenbrock's function has a global minimum reached in $(1,1)^\top$, which value is $f(1,1) = 0$.

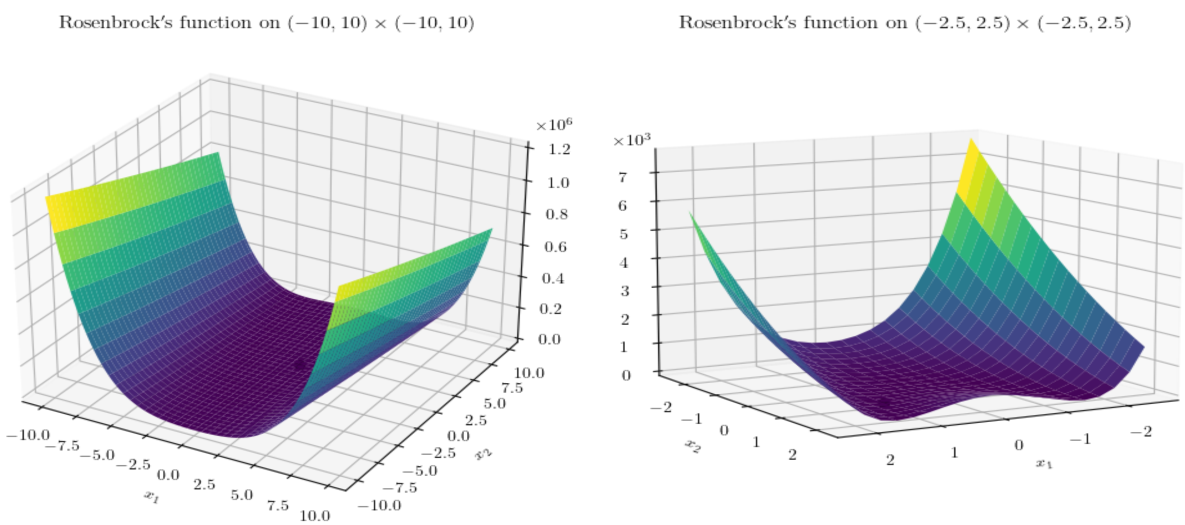

Figure 1: Rosenbrock's function

1.1 Constant stepsize

1.1.1 Starting point x^0

- Forall $t_k \geq 1$, $(x^k)_k$ oscillates in $[-2.1 \times 10^9, 2.1 \times 10^9] \times [-2.1 \times 10^9, 2.1 \times 10^9]$.
- Forall $0.0079 \leq t_k < 1$, $x^k \xrightarrow[k \to +\infty]{} (+\infty, +\infty)$.
- Let's study smaller setpsizes $t_k \leq 0.0078$. The Figure 2 shows the 2500 first iterations for $t_k = 0.002$. The red point is the starting point x^0 and the black point is the stationary point $(1, 1)$. It seems that $x^k \xrightarrow[k \to +\infty]{} (1, 1)$.

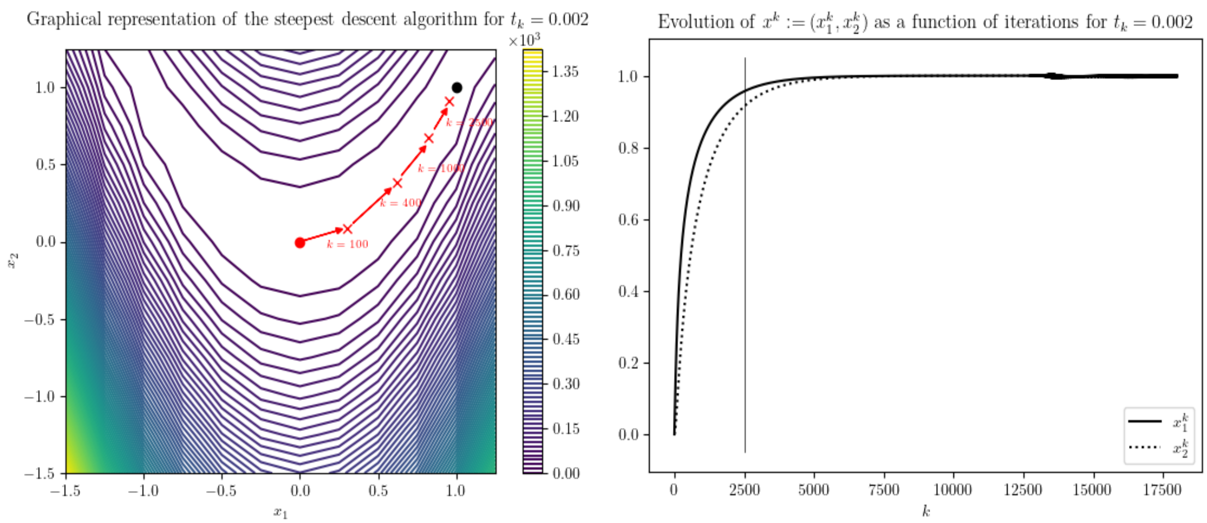

Figure 2: Steepest descent method using f and x^0 – focus on $(x^k)_{k \in \mathbb{N}}$

However the Figure 3, using several stepsizes, reveal that this is false. We see that, forall $t_k \leq 0.0078$, $\big(f(x^k)\big)_{k \in \mathbb{N}}$ converges by oscillation towards a positive value f_{t_k}. When t_k decreases, the limit value f_{t_k} decreases toward 0. Note : $f(x^0) = 1$.

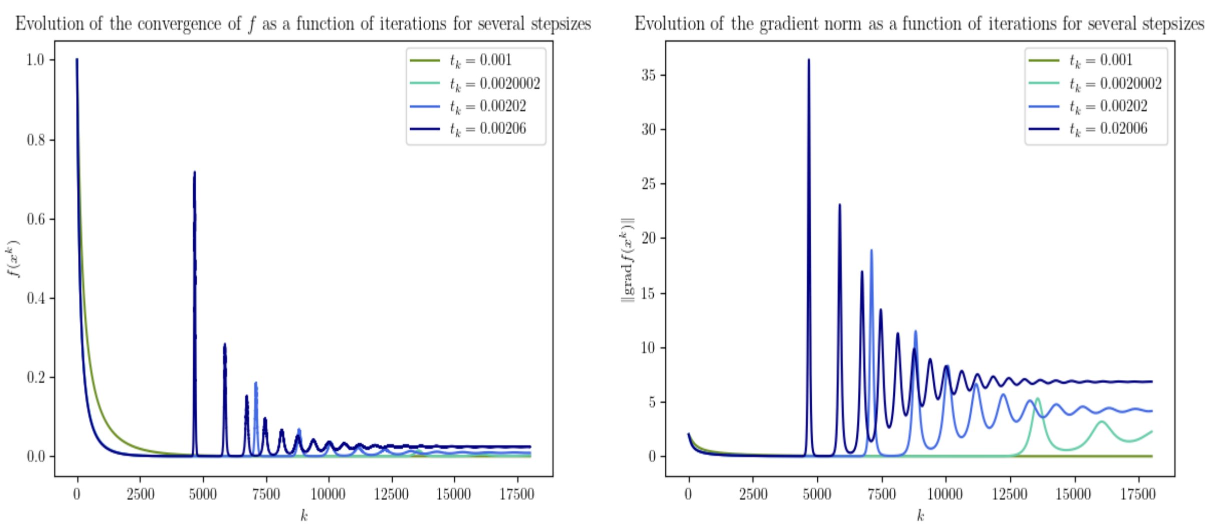

Figure 3: Steepest descent method using f and x^0 – focus on $\big(f(x^k)\big)_{k \in \mathbb{N}}$

Eventually, forall $t_k \in \mathbb{R}$, $\big(f(x^k)\big)_{k \in \mathbb{N}}$ is not converging to the minimum $f(1, 1) = 0$ of f.

1.1.2 Starting point $\tilde{x}^0$

- Forall $t_k \geqslant 0.00033779$, $x^k \xrightarrow[k \to +\infty]{} (\pm\infty, \pm\infty)$.
- Let's focus on smaller stepsizes $t_k \leq 0.0003377894$. On Figure 4 it seems that $f(x^k) \xrightarrow[k \to +\infty]{} 0$.

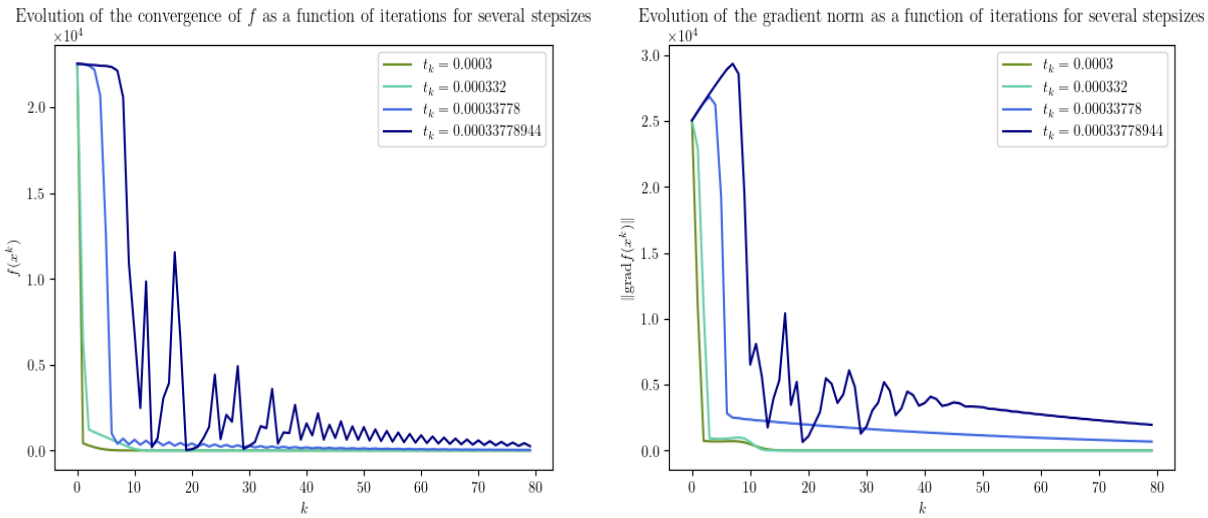

Figure 4: Steepest descent method using f and $\tilde{x}^0$ − focus on $\left(f(x^k)\right)_{k\in\mathbb{N}}$

However, we see on Figure 5 that, for instance, for $t_k = 0.000332$, $x^k \xrightarrow[k \to +\infty]{} (-1.68, 2.84) \neq (1, 1)$.

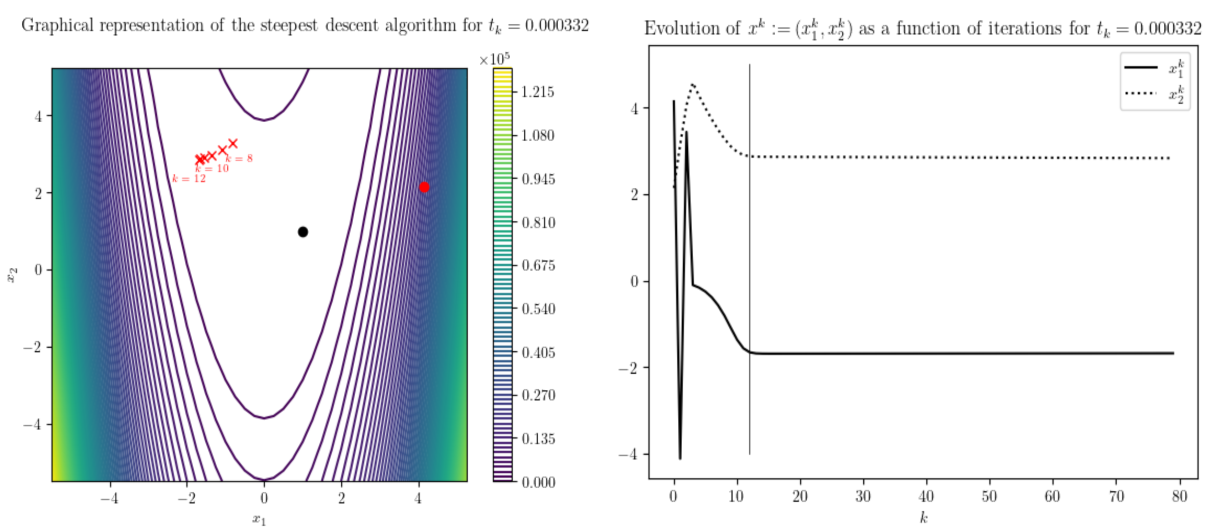

Figure 5: Steepest descent method using f and $\tilde{x}^0$ − focus on $(x^k)_{k\in\mathbb{N}}$

Eventually, forall $t_k \in \mathbb{R}$, $\left(f(x^k)\right)_{k\in\mathbb{N}}$ is not converging to the minimum $f(1,1) = 0$ of f.
Note : $f(\tilde{x}^0) \approx 2.3 \times 10^4$.

1.2 Nonconstant stepsize

1.2.1 Starting point x^0

For Armijo-stepsize, $x^k \xrightarrow[k\to+\infty]{} (1,1)$ so $f(x^k) \xrightarrow[k\to+\infty]{} f(1,1) = 0$ which is the global minimum of f.

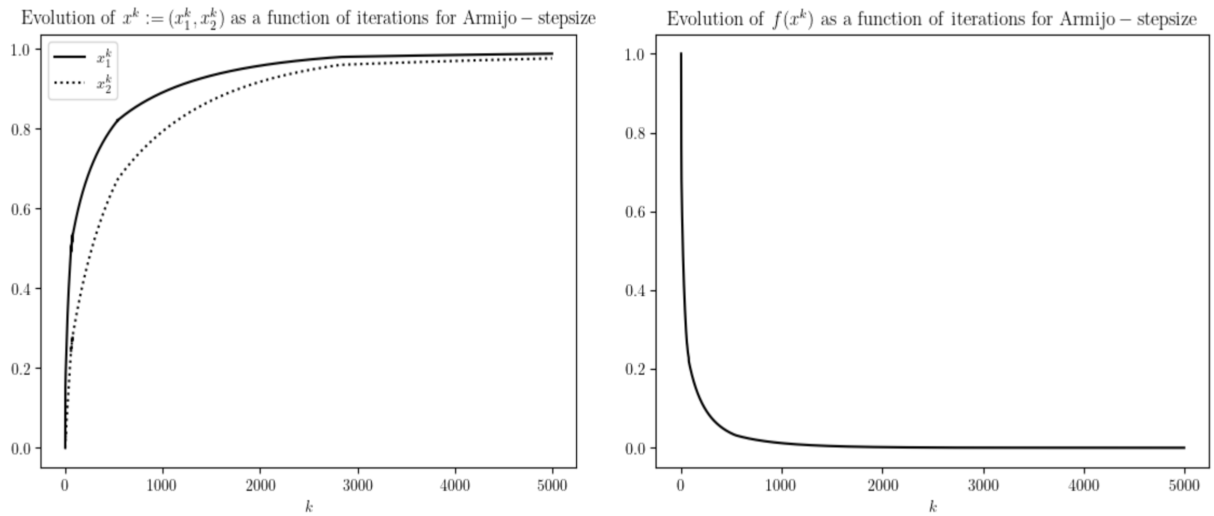

Figure 6: Steepest descent method using f, x^0 and Armijo-stepsize

1.2.2 Starting point $\tilde{x}^0$

For Armijo-stepsize, $x^k \xrightarrow[k\to+\infty]{} (1,1)$ so $f(x^k) \xrightarrow[k\to+\infty]{} f(1,1) = 0$ which is the global minimum of f.

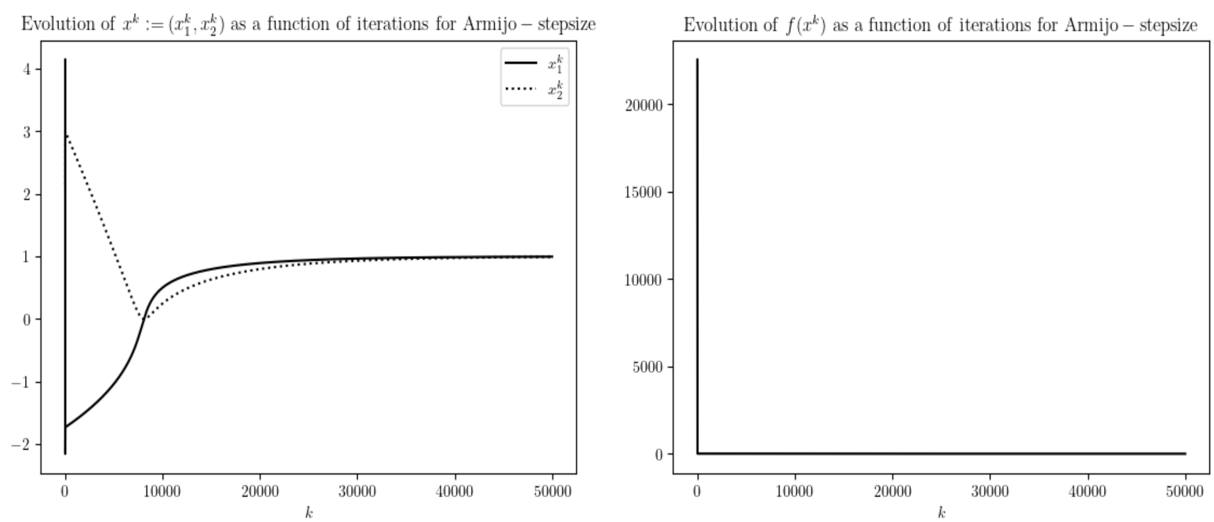

Figure 7: Steepest descent method using f, $\tilde{x}^0$ and Armijo-stepsize

2 Easom's function

Easom's function is differentiable.

$$
\begin{array}{rrcl}
g: & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\
 & (x_1, x_2) & \longmapsto & -\cos(x_1)\cos(x_2)\mathrm{e}^{-(x_1-\pi)^2-(x_2-\pi)^2}
\end{array}
$$

The gradient of g is given by

$$
\forall (x_1, x_2) \in \mathbb{R}^2 \quad \nabla g(x_1, x_2) = \begin{pmatrix} \cos(x_2)\mathrm{e}^{-(x_1-\pi)^2-(x_2-\pi)^2}\big(\sin(x_1) + 2\cos(x_1)(x_1 - \pi)\big) \\ \cos(x_1)\mathrm{e}^{-(x_1-\pi)^2-(x_2-\pi)^2}\big(\sin(x_2) + 2\cos(x_2)(x_2 - \pi)\big) \end{pmatrix}
$$

Easom's function has several local minima and a global minimum reached at $(\pi, \pi)^\top$ which value is $g(\pi, \pi) = -1$.

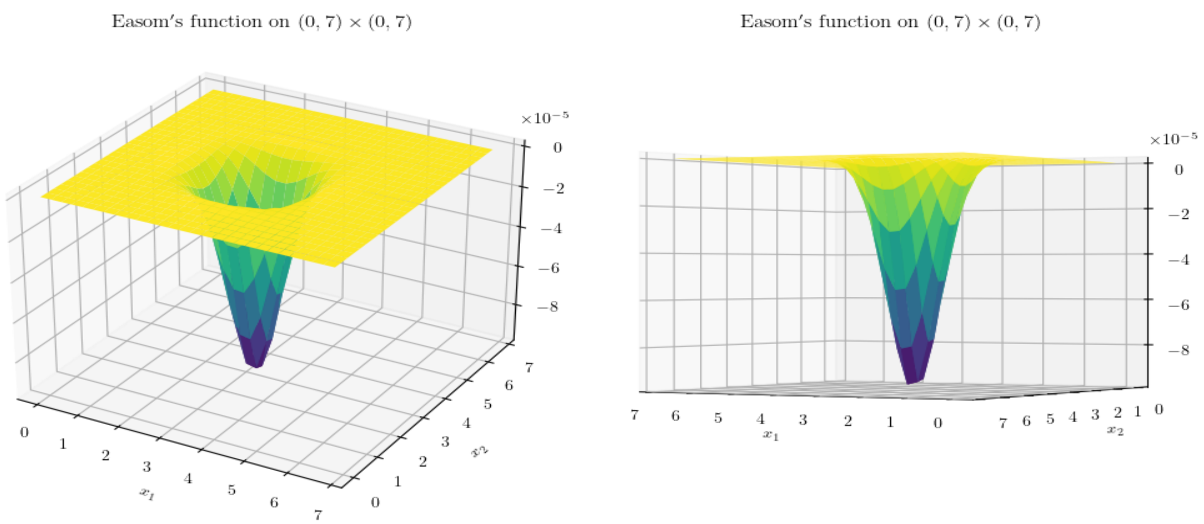

Figure 8: Easom's function

2.1 Constant stepsize

2.1.1 Starting point x^0

- Forall $t_k \geqslant 3955$ the algorithm converges to a point (x, x) depending on t_k on the *yellow plateau* (c.f. Figure 8) where $g(x, x) = 0$ is not the global minimum of g.

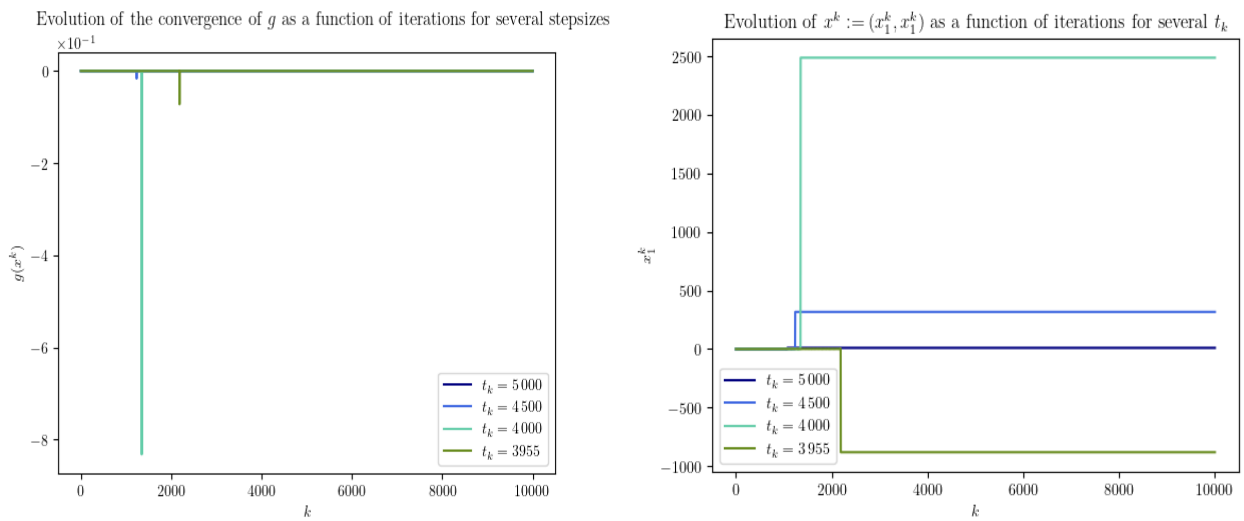

Figure 9: Steepest descent method using g and x^0

- Forall $t_k \leqslant 3954$, there exists a rank k_0 such that $(x^k)_{k \geqslant k_0}$ alternately takes two values (for instance, for $t_k = 1\,000$, $k_0 \approx 5200$). When t_k decreases, these two values (taken alternately) become close to $(1.30499545, 1.30499545)$ which is also on the *yellow plateau*. Obviously $\left(g(x^k)\right)_{k \geqslant k_0}$ also takes two values alternately. When t_k decreases these values become close to $g(1.30499545, 1.30499545) = -8.11022389 \times 10^{-5}$ which is not the global minimum of g.

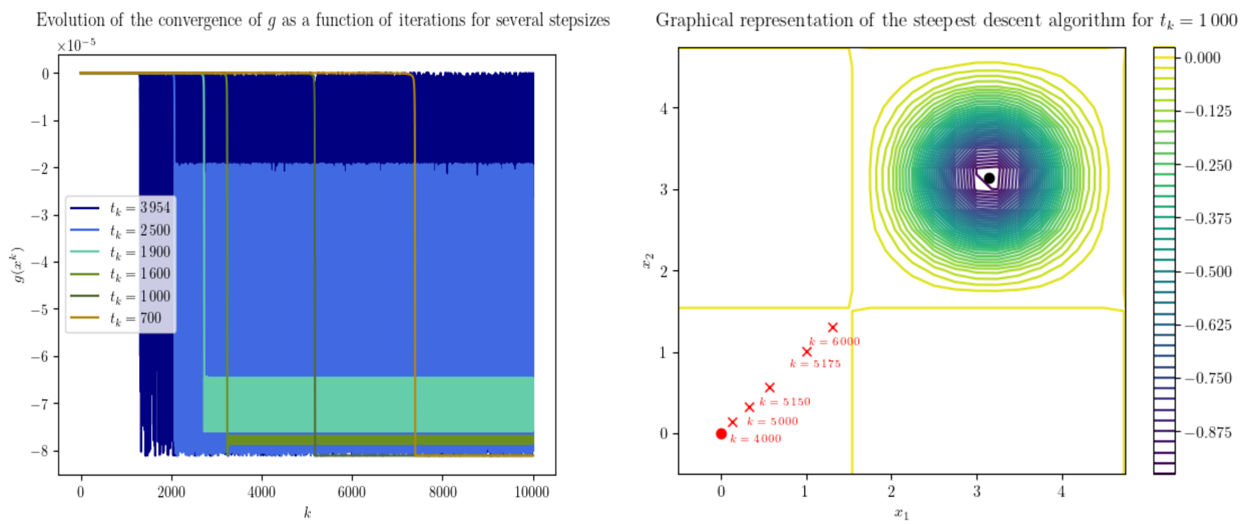

Figure 10: Steepest descent method using g and x^0

Eventually, forall $t_k \in \mathbb{R}$, $\left(g(x^k)\right)_{k \in \mathbb{N}}$ is not converging to the minimum $g(\pi, \pi) = -1$ of g.
Note : $g(x^0) \approx -2.68 \times 10^{-9}$.

2.1.2 Starting point $\tilde{x}^0$

- Forall $t_k \geq 3$, $x^k \xrightarrow[k\to+\infty]{} (x_1', x_2')$ where $g(x_1', x_2')$ is not the global minimum of g.

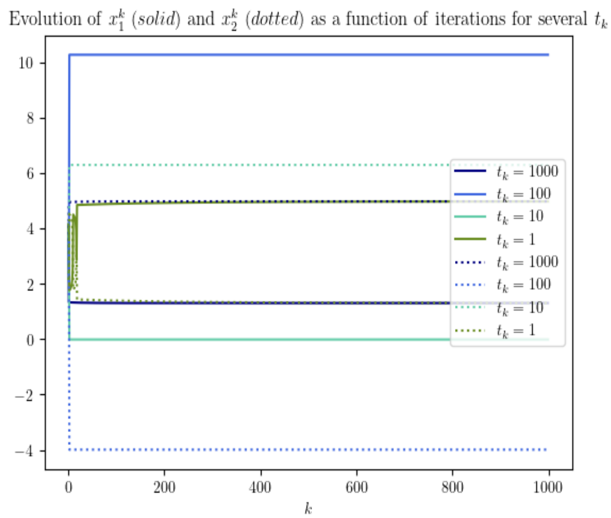

Figure 11: Steepest descent method using g and $\tilde{x}^0$

- For $0.6 < t_k < 3$, $(x^k)_{k\in\mathbb{N}}$ often oscillates.
- Forall $t_k \leq 0.6$, $x^k \xrightarrow[k\to+\infty]{} (\pi, \pi)$ so $g(x^k) \xrightarrow[k\to+\infty]{} -1$.

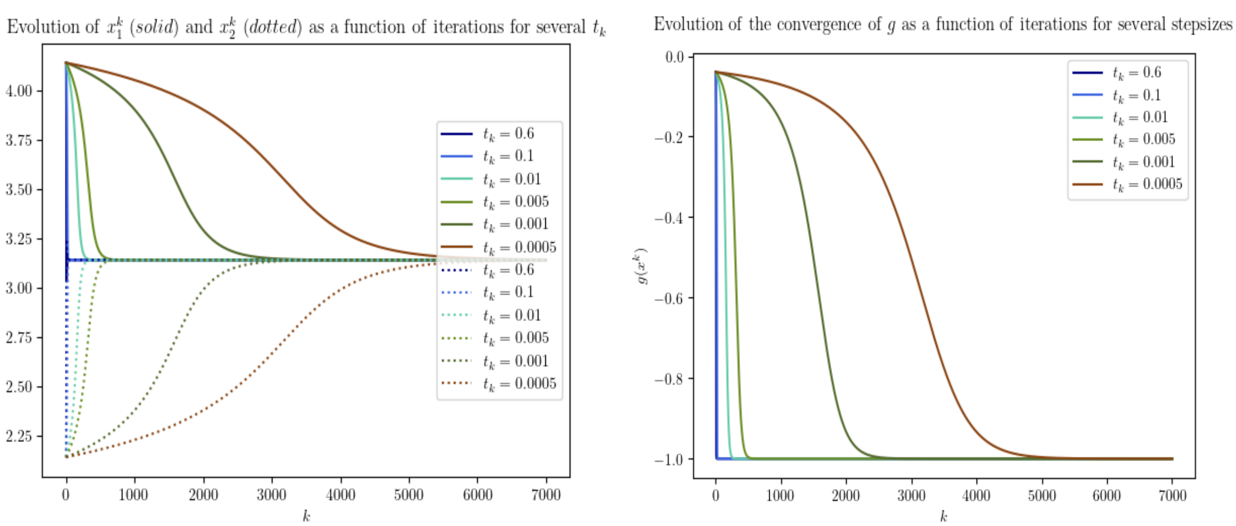

Figure 12: Steepest descent method using g and $\tilde{x}^0$

Eventually $\left(g(x^k)\right)_{k\in\mathbb{N}}$ converges toward the global minimum $g(\pi,\pi) = -1$ of g if and only if $t_k \leq 0.6$.
Note : $g(\tilde{x}^0) \approx -3.95 \times 10^{-2}$.

2.2 Nonconstant stepsize

2.2.1 Starting point x^0

For the Armijo-stepsize $(x^k)_{k\in\mathbb{N}}$ might converge toward (π, π) but very slowly. We should run the algorithm for much more than 50 000 iterations.

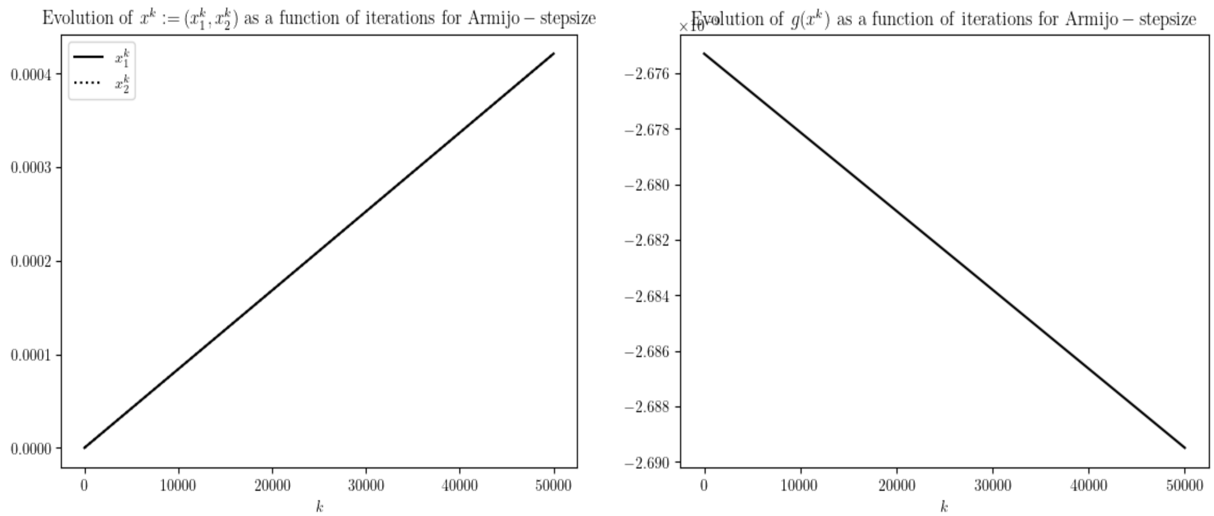

Figure 13: Steepest descent method using g, x^0 and Armijo-stepsize

2.2.2 Starting point $\tilde{x}^0$

For Armijo-stepsize, $x^k \xrightarrow[k\to+\infty]{} (\pi, \pi)$ so $g(x^k) \xrightarrow[k\to+\infty]{} g(\pi, \pi) = -1$ which is the global minimum of f. Figure 14 shows that this convergence is very fast.

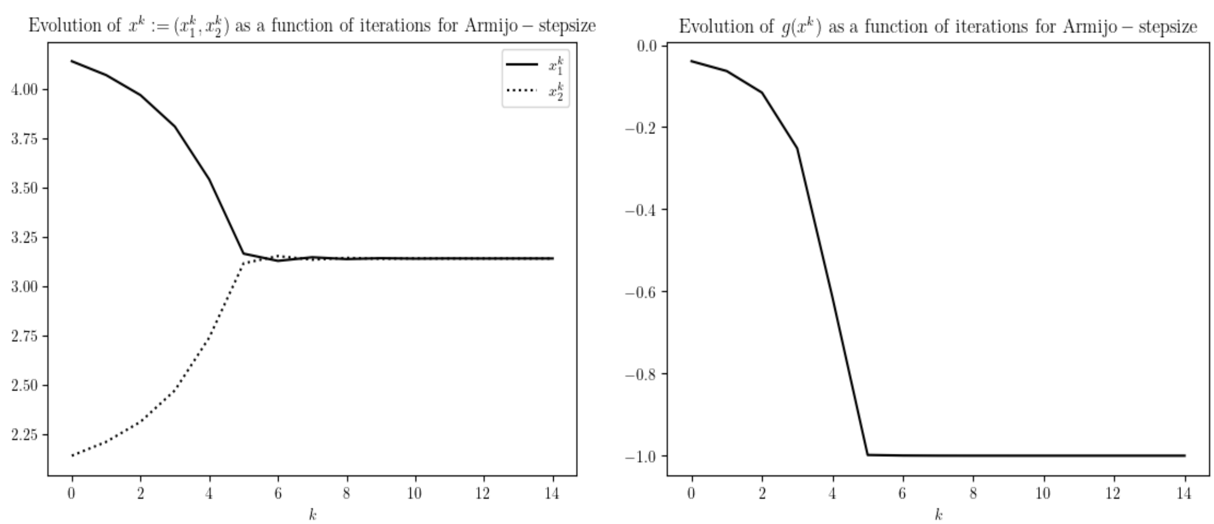

Figure 14: Steepest descent method using g, $\tilde{x}^0$ and Armijo-stepsize

3 Conclusion

Constant stepsizes are often not a good choice of stepsizes to study Rosenbrock's function and Easom's function with the steepest descent algorithm. Conversely, Armijo-stepsize lead to the convergence of the function toward his global minimum. However it should be mentioned that this convergence can be very slow. As one can see on Figure 15, Newton's algorithm converges much faster for f with starting point x^0 and $\tilde{x}^0$.

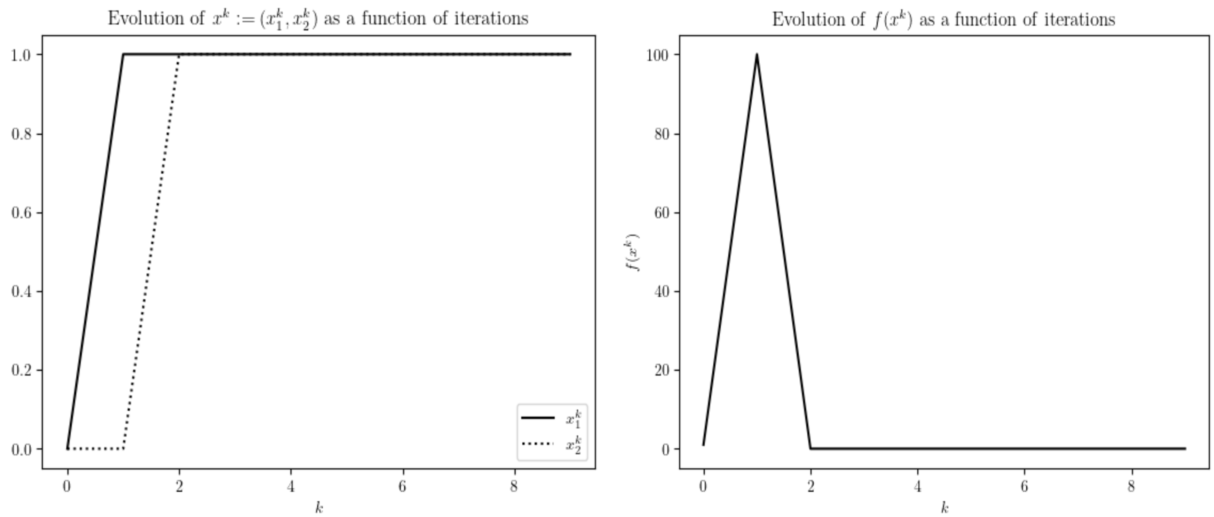

Figure 15: Newton's algorithm using f and x^0

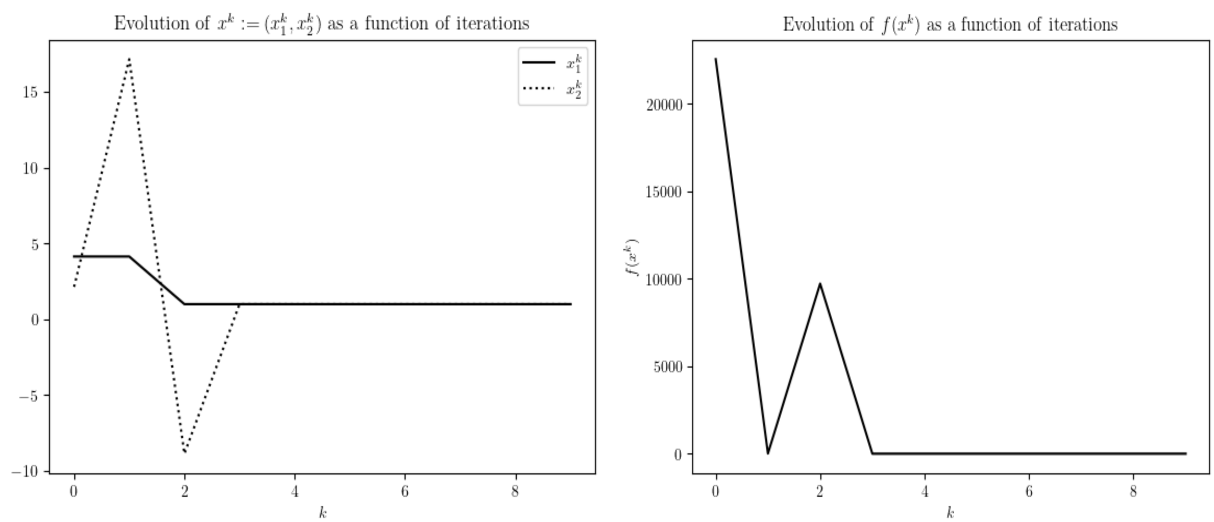

Figure 16: Newton's algorithm using f and $\tilde{x}^0$