



UNIVERSITÉ DE BORDEAUX

COLLÈGE SCIENCES ET TECHNOLOGIES

MASTER 2 MATHÉMATIQUES APPLIQUÉES ET STATISTIQUE — PARCOURS MSS

Algorithme de Nesterov et Équations Différentielles Ordinaires

*Projet d'Optimisation en grande dimension :
Approches déterministes et stochastiques*

Auteurs :

Théo Lahontâa
Elias Aboukacem
Zine Eddine Zidane
Mahamat Youssouf Souleyman

Encadrant : Camille Castera

Année universitaire : 2025–2026

Introduction

Le but de ce projet est d'étudier l'algorithme de Nesterov et son lien avec des équations différentielles ordinaires. On compare les dynamiques obtenues par discrétisation et on observe comment Nesterov accélère la descente de gradient. Les résultats théoriques et numériques nous permettront d'illustrer la différence entre les deux approches et le rôle du terme d'inertie.

1 Questions théoriques

1.1 Question 1

On considère l'EDO

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0, \quad t \geq 1, \quad x(1) = x_0, \quad \dot{x}(1) = 0,$$

avec f convexe, différentiable et minorée. En posant $v(t) = \dot{x}(t)$, on obtient le système du premier ordre :

$$\begin{aligned} \dot{x}(t) &= v(t), \\ \dot{v}(t) &= -\frac{\alpha}{t} v(t) - \nabla f(x(t)). \end{aligned}$$

On définit une grille temporelle régulière par

$$t_k = 1 + kh, \quad k = 0, 1, 2, \dots$$

afin de partir de la condition initiale au temps $t = 1$ et d'appliquer la discrétisation d'Euler avec un pas constant h . Ce choix rend l'évolution temporelle simple et cohérente, et permet d'évaluer le terme $\frac{\alpha}{t_k}$ à chaque itération.

Pour une EDO du type $\dot{z}(t) = F(t, z(t))$, le schéma d'Euler explicite est

$$\frac{z_{k+1} - z_k}{h} \approx \dot{z}(t_k) = F(t_k, z_k) \implies z_{k+1} = z_k + h F(t_k, z_k).$$

En l'appliquant au système précédent, on obtient en t_k :

$$x_{k+1} = x_k + h v_k, \quad v_{k+1} = v_k - h \frac{\alpha}{t_k} v_k - h \nabla f(x_k).$$

À l'indice $k - 1$, la première équation donne :

$$x_k = x_{k-1} + h v_{k-1} \implies v_{k-1} = \frac{x_k - x_{k-1}}{h}.$$

Au même indice $k - 1$, la seconde équation du schéma d'Euler s'écrit :

$$v_k = v_{k-1} - h \frac{\alpha}{t_{k-1}} v_{k-1} - h \nabla f(x_{k-1}) = \left(1 - \frac{\alpha h}{t_{k-1}}\right) v_{k-1} - h \nabla f(x_{k-1}).$$

En remplaçant v_{k-1} puis en injectant dans $x_{k+1} = x_k + h v_k$, on obtient :

$$x_{k+1} = x_k + \left(1 - \frac{\alpha h}{t_{k-1}}\right) (x_k - x_{k-1}) - h^2 \nabla f(x_{k-1}).$$

En posant $h = 1$ et $t_k = 1 + k$, cela devient :

$$x_{k+1} = x_k + \frac{k - \alpha}{k} (x_k - x_{k-1}) - \gamma \nabla f(x_{k-1}), \quad \gamma := h^2 = 1.$$

L'algorithme de Nesterov est défini par :

$$y_k = x_k + \frac{k + 1 - \alpha}{k + 1} (x_k - x_{k-1}), \quad x_{k+1} = y_k - \gamma \nabla f(y_k).$$

À l'indice $k+1$, le schéma d'Euler explicite est différent de Nesterov sur 2 points principaux :

- Le gradient est évalué en x_{k-1} , alors que Nesterov l'évalue en le point extrapolé y_k ;
- le coefficient d'inertie est $\frac{k-\alpha}{k}$ (avec $h = 1$), tandis que Nesterov utilise $\frac{k+1-\alpha}{k+1}$.

Ces différences montrent que la discrétisation d'Euler de $\ddot{x} + \frac{\alpha}{t} \dot{x} + \nabla f(x) = 0$ ne reproduit pas exactement l'algorithme de Nesterov.

1.2 Question 2

Pour l'EDO :

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t) + \beta \dot{x}(t)) = 0,$$

On suit la même démarche que dans la Q1. En appliquant Euler explicite au système du premier ordre associé, puis en éliminant v_k comme dans la question précédente, on obtient :

$$x_{k+1} = x_k + \left(1 - \frac{\alpha h}{t_{k-1}}\right) (x_k - x_{k-1}) - h^2 \nabla f\left(x_{k-1} + \frac{\beta}{h} (x_k - x_{k-1})\right).$$

On note :

$$a_{k-1} := 1 - \frac{\alpha h}{t_{k-1}}, \quad y_{k-1} := x_{k-1} + a_{k-1} (x_k - x_{k-1}).$$

Pour que le gradient soit évalué au même point extrapolé, on impose l'égalité :

$$x_{k-1} + \frac{\beta}{h} (x_k - x_{k-1}) = x_{k-1} + a_{k-1} (x_k - x_{k-1}), \quad \Rightarrow \quad \frac{\beta}{h} = a_{k-1}.$$

On obtient alors :

$$x_{k+1} = y_{k-1} - h^2 \nabla f(y_{k-1}).$$

Avec $h = 1$ et $t_k = 1 + k$, on a :

$$a_k = \frac{k+1-\alpha}{k+1}, \quad \beta_k = h a_k = 1 - \frac{\alpha}{k+1}.$$

En réécrivant la relation au rang k :

$$y_k = x_k + \frac{k+1-\alpha}{k+1}(x_k - x_{k-1}), \quad x_{k+1} = y_k - \gamma \nabla f(y_k), \quad \gamma = h^2,$$

ce qui correspond bien à l'algorithme de Nesterov.

2 Implémentation numérique

2.1 Question 3

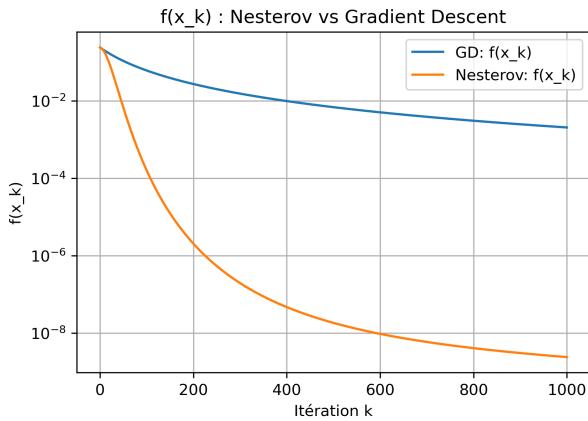


FIGURE 1 – Courbe de $f(x_k)$ pour l'algorithme de Nesterov et Gradient Descent sur la fonction $f(x) = \frac{1}{4}x^4$.

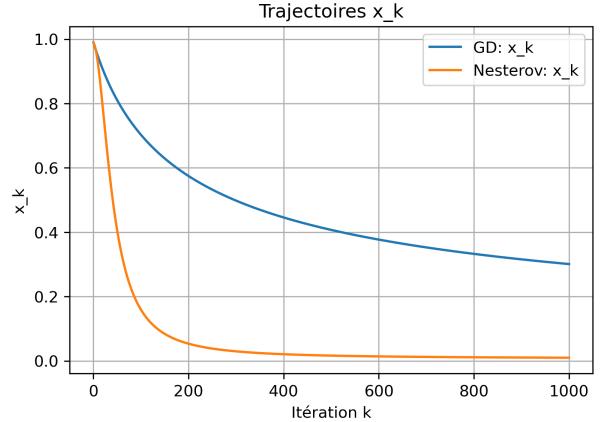


FIGURE 2 – Trajectoire des itérés x_k pour l'algorithme de Nesterov et Gradient Descent.

Sur la Figure 1 ,on observe que Nesterov converge plus vite que la descente de gradient. Après environ 200 itérations, $f(x_k)$ atteint 10^{-6} avec Nesterov, alors que la descente de gradient reste autour de 10^{-2} . La Figure 2 illustre la trajectoire de x_k , on voit aussi que x_k se rapproche de 0 plus rapidement avec Nesterov.

2.2 Question 4 et 5

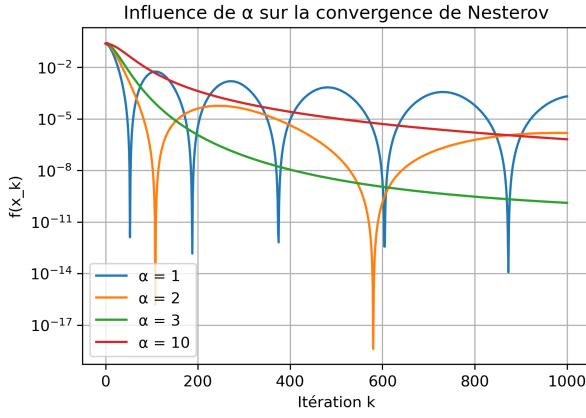


FIGURE 3 – Courbe de $f(x_k)$ pour différentes valeurs de α dans l’algorithme de Nesterov.

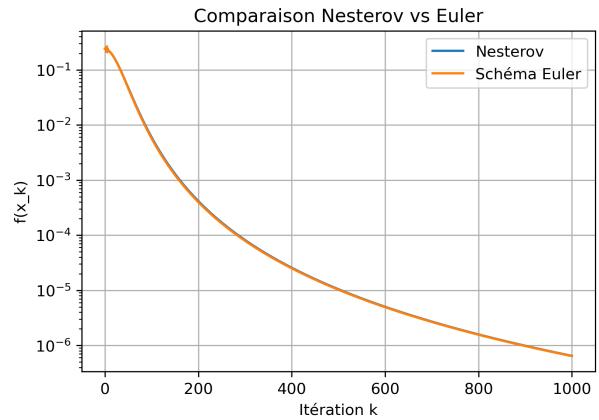


FIGURE 4 – Comparaison de $f(x_k)$ entre le schéma d’Euler explicite et Nesterov.

Pour $\alpha = 1$ et $\alpha = 2$, la dynamique présente des fortes oscillations : $f(x_k)$ atteint ponctuellement des valeurs très petites, mais l’évolution n’est pas régulière et ne montre pas une décroissance stable. Pour $\alpha = 10$, $f(x_k)$ décroît sans oscillations mais lentement. Dans ce cas, l’inertie est trop faible, ce qui limite l’effet d’accélération caractéristique de Nesterov. Afin d’obtenir un compromis entre stabilité et rapidité, une valeur intermédiaire $\alpha = 3$ a été testée. Cette valeur supprime les oscillations observées pour $\alpha = 1$ et $\alpha = 2$, tout en assurant une décroissance sensiblement plus rapide que pour $\alpha = 10$.

Ainsi, $\alpha \approx 3$ apparaît comme un choix particulièrement pertinent, offrant un bon équilibre entre stabilité et vitesse de convergence.

Les deux méthodes donnent des courbes superposées : pour $f(x) = \frac{1}{4}x^4$, l’évaluation du gradient en y_k (Nesterov) ou en x_k (Euler) produit une même vitesse de convergence. L’écart est négligeable car $x_k \approx y_k$ dès les premières itérations, mais sur des fonctions plus complexes, cela pourrait être différent.

2.3 Question 6

Pour les fonctions convexes, la descente de gradient converge en

$$f(x_k) - f(x^*) = O\left(\frac{1}{k}\right),$$

tandis que l’algorithme de Nesterov converge en

$$f(x_k) - f(x^*) = O\left(\frac{1}{k^2}\right).$$

Nesterov converge donc plus rapidement que la descente de gradient.

2.4 Question 7

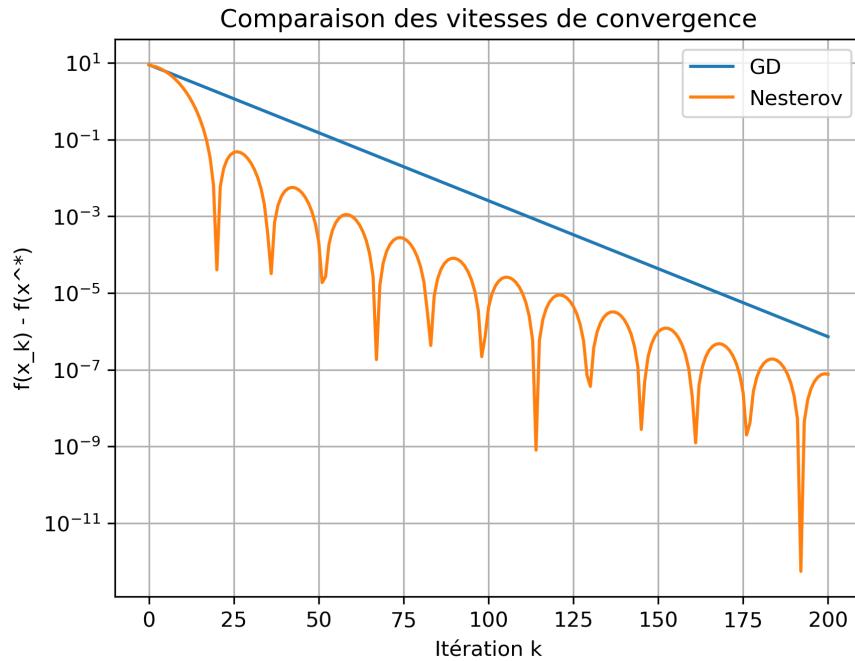


FIGURE 5 – Comparaison de des vitesses de convergence entre GD et Nesterov.

La courbe de la figure 5 montre que la décroissance est plus rapide que celle en $1/k$ malgré les oscillations. La vitesse observée semble donc cohérente avec un comportement en $1/k^2$.

2.5 Question 8

La convergence est plus rapide car la fonction utilisée ici est une fonction quadratique strictement convexe dont la matrice associée est mieux conditionnée que dans les expériences précédentes. On le voit au fait que la descente de gradient elle-même devient plus rapide. D'après le cours, un meilleur conditionnement accélère les deux méthodes, et Nesterov en profite également, ce qui explique la vitesse observée.