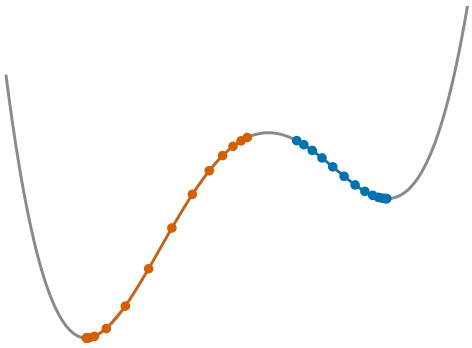
Équations différentielles et optimisation
L3 Mathématiques · Travaux pratiques

La méthode du gradient

Descendre une pente, pas à pas

Introduction au TP 7



Santiago Toro Oquendo
Université d'Artois · Semestre 1, 2026–2027

Du TP 6 au TP 7 : trouver le minimum

Au TP 6, on a *reconnu* les minima : ce sont des points critiques, où la dérivée (le gradient) s'annule, et la dérivée seconde (la hessienne) dit s'il s'agit d'un minimum.

Mais, sauf dans les cas simples, on ne sait pas résoudre $f'(a) = 0$ à la main.

Au TP 7, on *cherche* le minimum avec un algorithme : partir d'un point et descendre, pas à pas,

$$a_{k+1} = a_k - e\, f'(a_k) \qquad (e > 0 : \text{le pas}).$$

Puis on s'en sert pour une **régression linéaire**, et on lui donne de l'**élan** : le moment.

Un peu d'histoire

Cauchy propose la méthode en 1847 pour des calculs d'astronomie : « Méthode générale pour la résolution des systèmes d'équations simultanées », *C. R. Acad. Sci. Paris* 25, p. 536–538.

C'est encore elle, sous le nom de *rétropropagation*, qui entraîne les réseaux de neurones.

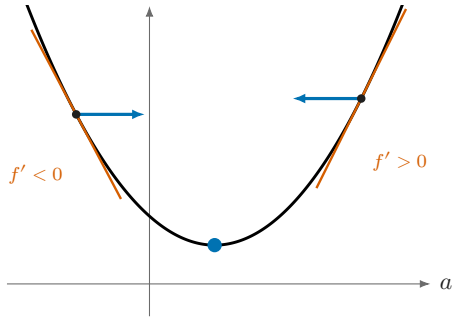
L'idée : descendre la pente

Le randonneur dans le brouillard. Il ne voit que la pente sous ses pieds : il fait un pas dans le sens où ça descend, puis il recommence.

En une variable, la pente est la **dérivée** :

- $f'(a) > 0$: f croît près de a, il faut aller vers la **gauche** ;
- $f'(a) < 0$: f décroît près de a, il faut aller vers la **droite**.

> **Règle**
>
> Pour faire baisser f, on se déplace dans le sens **opposé** au signe de $f'(a)$.



en orange, la tangente ; en bleu, le sens de la descente

La dérivée : la pente de la tangente

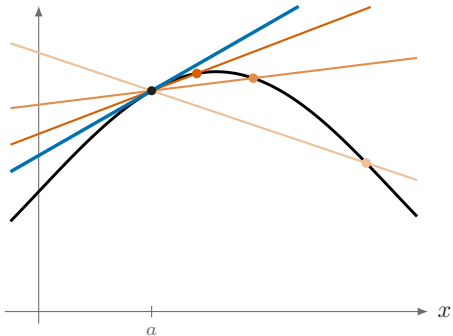
La **sécante** passe par $(a, f(a))$ et $(a+h, f(a+h))$; sa pente est le taux d'accroissement. Quand $h \to 0$, elle tend vers la **dérivée** :

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

La **tangente** $y = f(a) + f'(a)\,(x - a)$ est la droite qui approche le mieux la courbe près de a :

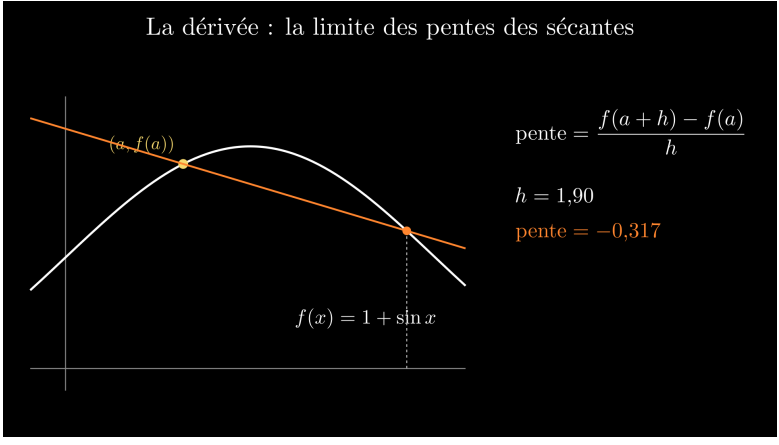
$$f(a+h) = f(a) + f'(a)\,h + o(h).$$

C'est la formule de Taylor à l'ordre 1 : de près, la courbe ressemble à sa tangente.



$f(x) = 1 + \sin x$, $a = 1$: sécantes pour $h = 1{,}9\,;\ 0{,}9\,;\ 0{,}4$ (orange) et tangente (bleu)

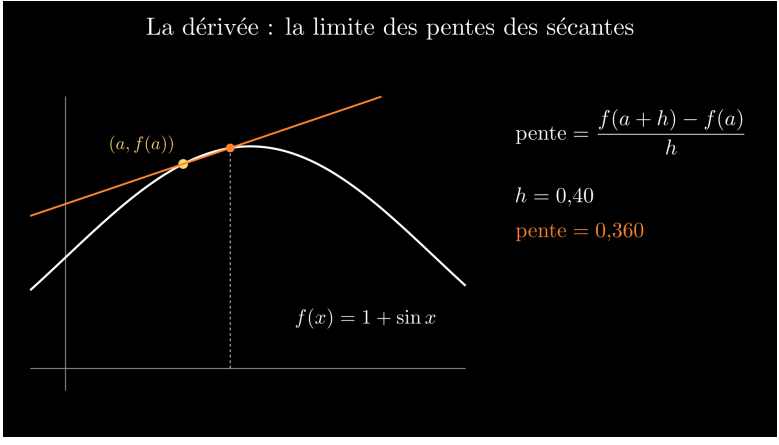
La dérivée, en mouvement



La sécante entre $(a, f(a))$ et $(a+h, f(a+h))$, pour $h = 1{,}9$: sa pente vaut $-0{,}317$.

image 1 / 3 · ▶ la vidéo (17 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

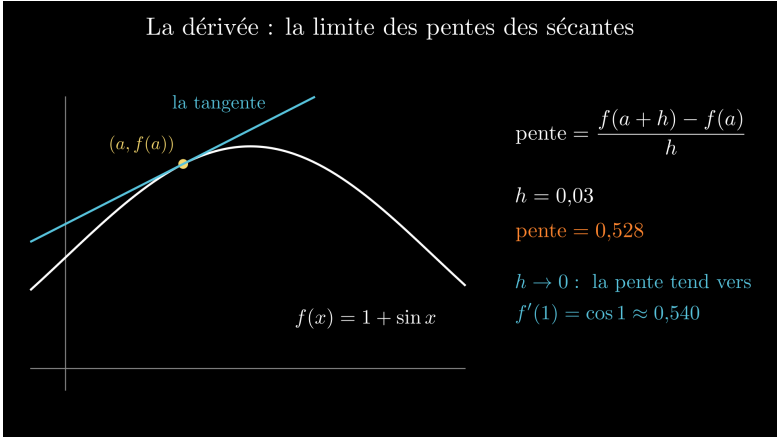
La dérivée, en mouvement



$h = 0{,}4$: la sécante a tourné autour de $(a, f(a))$; sa pente vaut $0{,}360$.

image 2 / 3 · ▶ la vidéo (17 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

La dérivée, en mouvement



$h \to 0$: les sécantes tendent vers la tangente, de pente $f'(1) = \cos 1 \approx 0{,}540$.

image 3 / 3 · ▶ la vidéo (17 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

La dérivée numérique (rappel du TP 6)

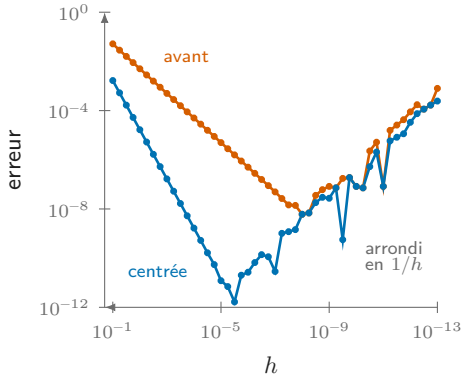
On remplace la limite par un petit h. D'après Taylor :

- différence **avant** : $\dfrac{f(a+h)-f(a)}{h} = f'(a) + \dfrac{f''(a)}{2}\,h + \dots$, erreur en h ;
- différence **centrée** : $\dfrac{f(a+h)-f(a-h)}{2h} = f'(a) + \dfrac{f'''(a)}{6}\,h^2 + \dots$, erreur en h^2.

Mais f est calculée avec environ 16 chiffres : l'arrondi ajoute une erreur de l'ordre de $10^{-16}/h$. Il ne faut donc pas prendre h trop petit.

Au TP

La différence centrée, avec $h = 10^{-5}$: une dizaine de chiffres exacts.



erreur sur $f'(0)$ pour $f = \exp$; les meilleurs pas sont $h \approx 10^{-8}$ (avant) et $h \approx 10^{-5}$ (centrée)

Deux idées pour descendre

Le gradient « brutal » : des pas de longueur *fixe* e, dans le bon sens,

$$a_{k+1} = a_k - \text{signe}\left(f'(a_k)\right) e.$$

Près du minimum, il fait des allers-retours sans jamais s'arrêter.

La méthode du gradient : des pas *proportionnels à la pente*,

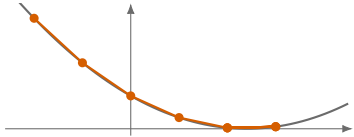
$$a_{k+1} = a_k - e\, f'(a_k).$$

Les pas raccourcissent d'eux-mêmes près du minimum, où $f' \approx 0$.

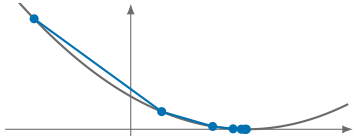
Pourquoi f diminue

$f(a - e\, f'(a)) = f(a) - e\, f'(a)^2 + o(e)$ (Taylor).
Si la suite converge vers a^*, alors $f'(a^*) = 0$.

pas fixe $e = 0{,}5$: allers-retours



méthode du gradient, $e = 0{,}6$



$q(a) = \frac{1}{2}(a - 1{,}2)^2$, depuis $a_0 = -1$

Le rôle du pas : un exemple en dimension 1

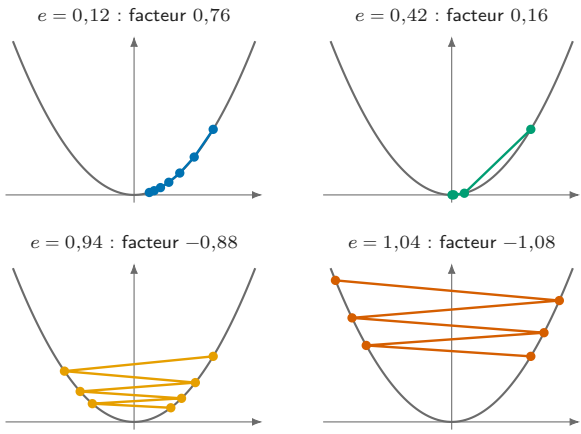
Pour $f(x) = x^2$, $f'(x) = 2x$, donc

$$x_{k+1} = x_k - 2e\,x_k = (1 - 2e)\,x_k.$$

À chaque pas, x_k est multiplié par le **facteur** $1 - 2e$:

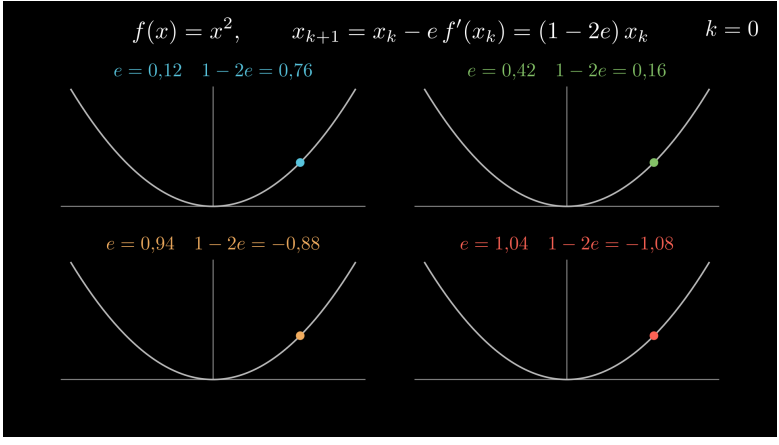
- entre 0 et 1 : convergence, d'autant plus lente que e est petit ;
- entre -1 et 0 : les itérés **changent de signe** (oscillations), mais convergent ;
- moins que -1, c'est-à-dire $e > 1$: $|x_k| \to \infty$.

Le seuil $e = 1$ vaut $2/f''$: plus f est courbée, plus le pas doit être petit.



itérés $(x_k, f(x_k))$ depuis $x_0 = 1{,}6$: lent, rapide, oscillant, divergent

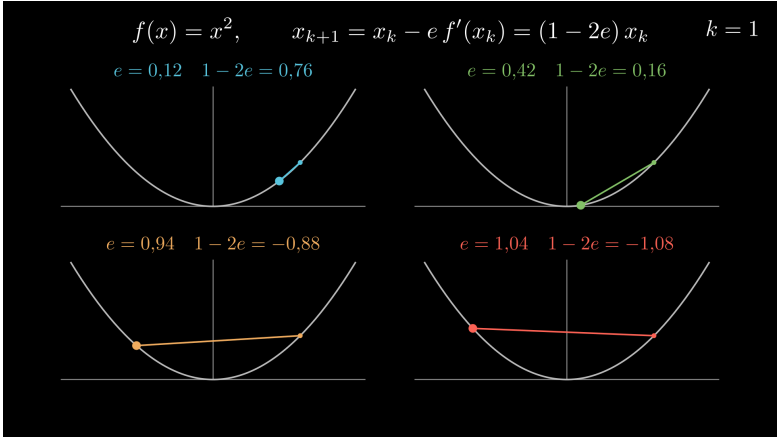
Le rôle du pas, en mouvement



Quatre pas e, même départ $x_0 = 1{,}6$ sur $f(x) = x^2$.

image 1 / 4 · ▶ la vidéo (17 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

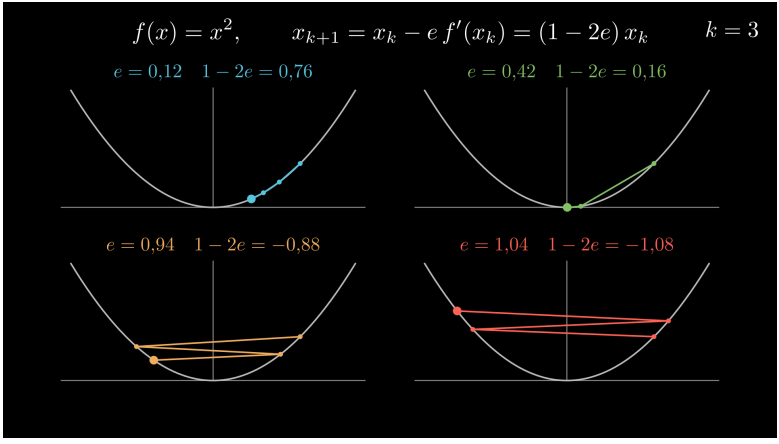
Le rôle du pas, en mouvement



Une itération : chaque point est multiplié par son facteur $1-2e$.

image 2 / 4 · ▶ la vidéo (17 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

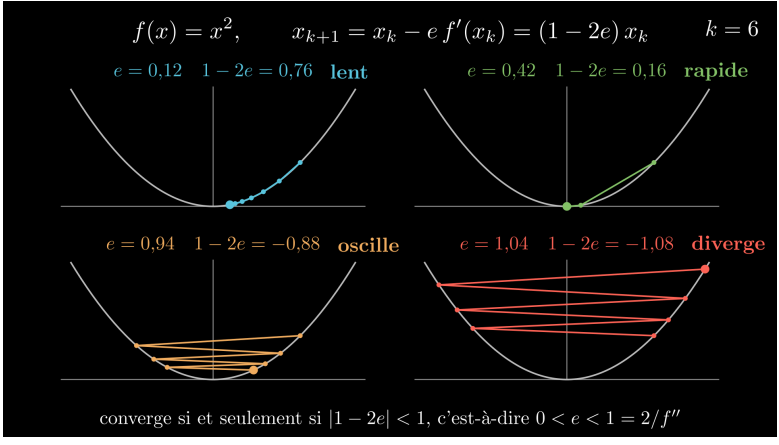
Le rôle du pas, en mouvement



Trois itérations : $e = 0{,}42$ est presque au minimum ; $e = 0{,}94$ change de côté à chaque pas.

image 3 / 4 · ▶ la vidéo (17 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

Le rôle du pas, en mouvement



Six itérations : lent, rapide, oscillant, divergent. Convergence si et seulement si $0 < e < 1 = 2/f''$.

image 4 / 4 · ▶ la vidéo (17 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

Près d'un minimum : la courbure fixe le pas

Près d'un minimum a^*, Taylor appliqué à f' donne $f'(a) \approx f''(a^*)\,(a - a^*)$, donc

$$a_{k+1} - a^* \approx \big(1 - e\, f''(a^*)\big)\,(a_k - a^*) :$$

c'est l'exemple de x^2, avec $f''(a^*)$ à la place de 2.

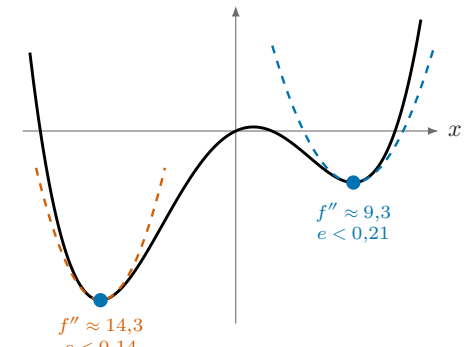
Le seuil

La méthode converge vers a^* si

$$0 < e < \frac{2}{f''(a^*)} :$$

plus la cuvette est **étroite** ($f''(a^*)$ grand), plus le pas doit être **petit**.

Au-delà du seuil, les points s'écartent de a^* en oscillant.

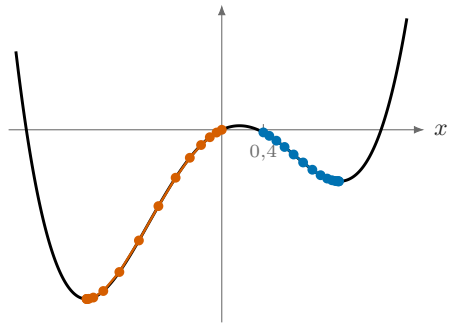


$g(x) = x^4 - 3x^2 + x$ (TP 6) et, en pointillés, ses paraboles osculatrices aux deux minima

Plusieurs minima : le départ compte

Si f a plusieurs minima locaux, la méthode du gradient va vers l'un d'eux : **lequel dépend du point de départ**.

- Elle descend dans la cuvette où elle se trouve : elle ne « voit » pas le reste de la courbe.
- Le minimum trouvé est **local**, pas forcément le plus bas.
- Elle s'arrête en un point critique ($f' = 0$) : en général un minimum local.



même g, pas $0{,}05$: depuis 0 (orange), on va dans la cuvette de gauche ; depuis $0{,}4$ (bleu), dans celle de droite

Application : la régression linéaire

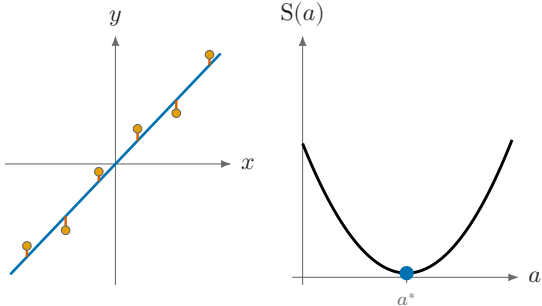
Des points (x_i, y_i) ; on cherche la droite $y = a\,x$ qui les approche le mieux.

- Le **résidu** $y_i - a\,x_i$: l'écart vertical entre le point et la droite (en orange).
- La **méthode des moindres carrés** choisit la pente qui minimise l'action

$$\mathrm{S}(a) = \sum_i (y_i - a\,x_i)^2.$$

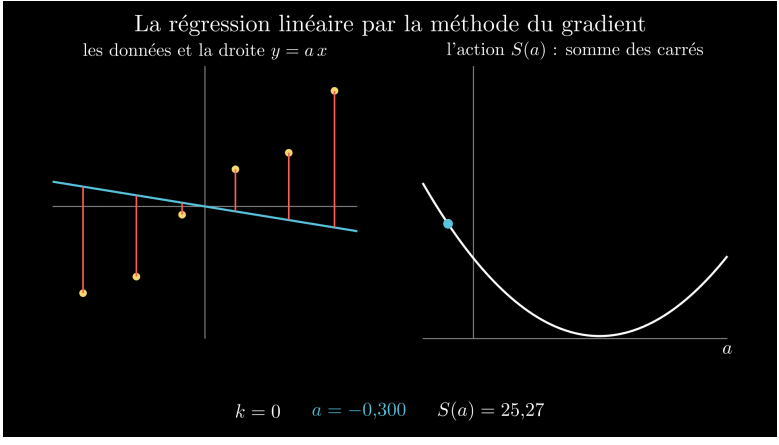
- S est un polynôme de degré 2 en a : une parabole, qui a un seul minimum.

Au TP : la méthode du gradient appliquée à S, comparée à la formule exacte, que vous trouverez.



à gauche, six points, la droite $y = a^* x$ et les résidus ; à droite, l'action S, minimale en $a^* \approx 1{,}49$

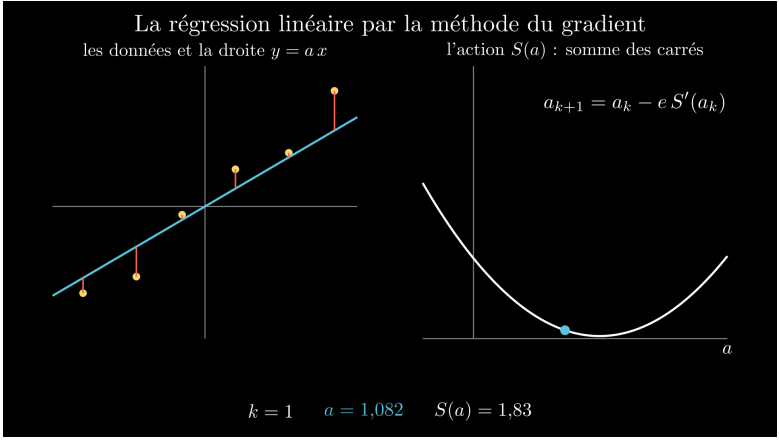
La régression, en mouvement



Une droite au hasard ($a_0 = -0{,}3$) : de grands résidus (en rouge), et $\mathrm{S}(a_0) \approx 25{,}3$.

image 1 / 4 · ▶ la vidéo (18 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

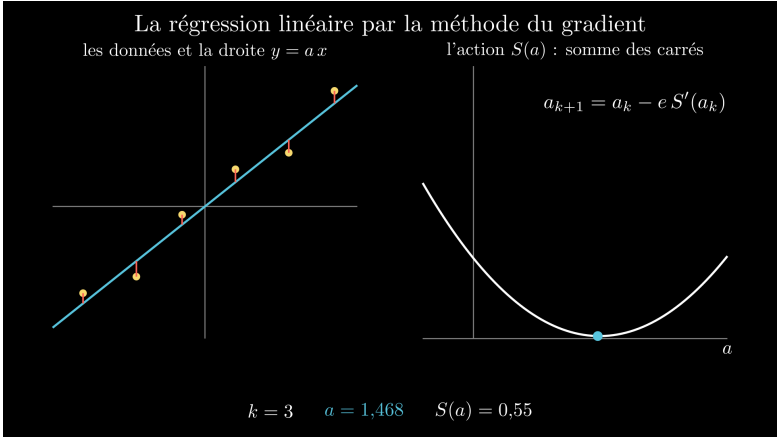
La régression, en mouvement



Une itération, $a_1 = a_0 - e\,\mathrm{S}'(a_0) \approx 1{,}082$: la droite tourne, et S tombe à $1{,}83$.

image 2 / 4 · ▶ la vidéo (18 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

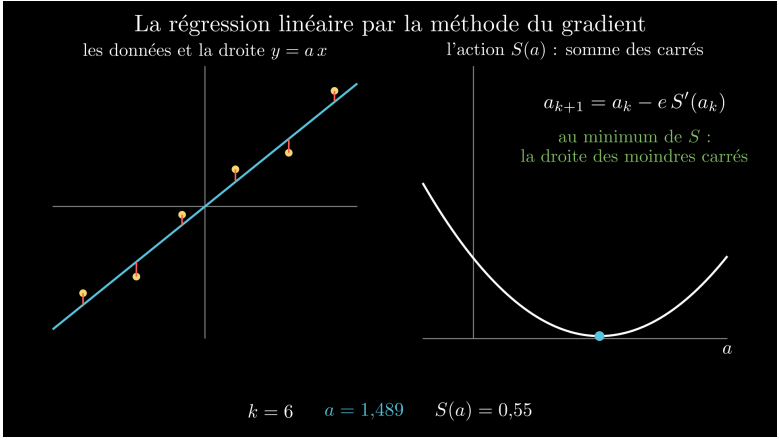
La régression, en mouvement



Trois itérations : $a \approx 1{,}468$; les pas raccourcissent, car $\mathrm{S}'(a)$ tend vers 0.

image 3 / 4 · ▶ la vidéo (18 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

La régression, en mouvement



Au minimum de S ($a \approx 1{,}489$, S $\approx 0{,}55$) : la droite des moindres carrés.

image 4 / 4 · ▶ la vidéo (18 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

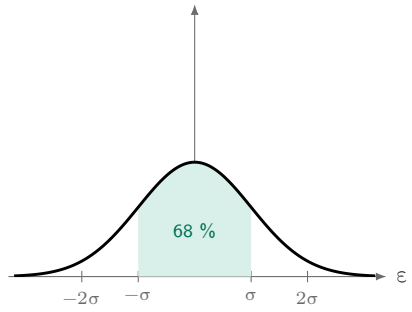
Fabriquer des données : le bruit gaussien

Pour tester une méthode, on fabrique des points autour d'une droite connue :

$$y_i = a\,x_i + b + \varepsilon_i.$$

- Le **bruit** ε_i suit une loi normale (gaussienne) de moyenne 0 et d'écart-type σ : environ 68 % des tirages entre $-\sigma$ et σ, et 95 % entre -2σ et 2σ.
- En Python : `rng = np.random.default_rng(graine)`, puis `rng.normal(m, s, n)`.
- Même graine, mêmes tirages : l'expérience est **reproductible**.

La pente retrouvée est proche de a, sans lui être égale : le bruit compte.



densité de la loi normale de moyenne 0 et d'écart-type σ

Le moment : donner de l'élan

Une boule qui roule garde son élan : elle traverse les zones plates et franchit les petites bosses.

Le moment (« boule pesante », Polyak, 1964)

$$\mathrm{M}_0 = 0, \qquad \mathrm{M}_{k+1} = b\,\mathrm{M}_k + e\,f'(a_k),$$
$$a_{k+1} = a_k - \mathrm{M}_{k+1},$$

avec $0 \leqslant b < 1$: M garde la mémoire des pas précédents.

- $b = 0$: c'est la méthode du gradient ; b proche de 1 : les pas qui vont dans le même sens s'accumulent.
- Avec trop d'élan, la boule dépasse le minimum et oscille plus longtemps avant de s'arrêter.

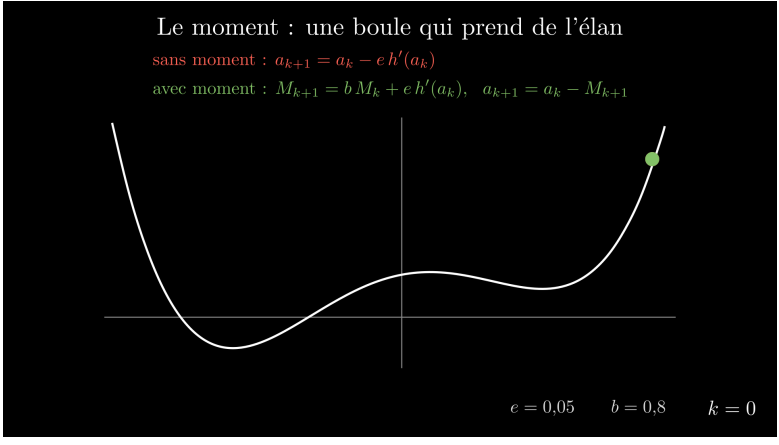
Le lien avec les EDO

- La méthode du gradient est la méthode d'**Euler** (TP 2), de pas e, pour $x'(s) = -f'(x(s))$: on suit la pente, sans inertie.
- Le moment discrétise l'équation d'une **boule pesante avec frottement**,

$$x'' + \gamma\, x' + f'(x) = 0,$$

avec un pas de temps τ : $b = 1 - \gamma\tau$ et $e = \tau^2$.

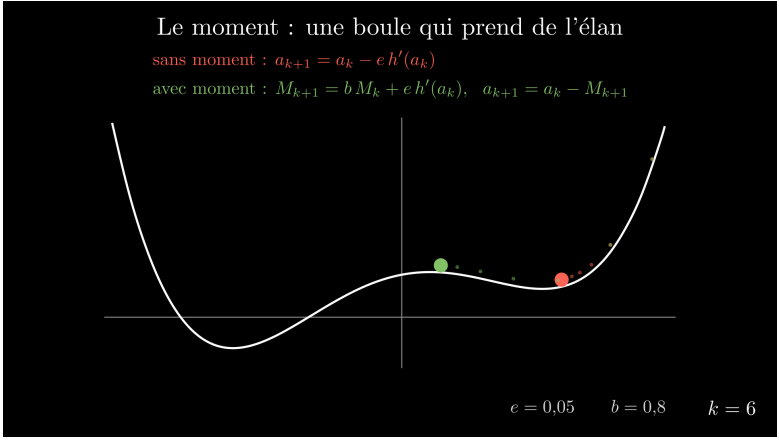
Le moment, en mouvement



Même départ ($a_0 = 1{,}6$), même pas ($e = 0{,}05$) ; la boule verte, devant la rouge, a le moment $b = 0{,}8$.

image 1 / 4 · ▶ la vidéo (21 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

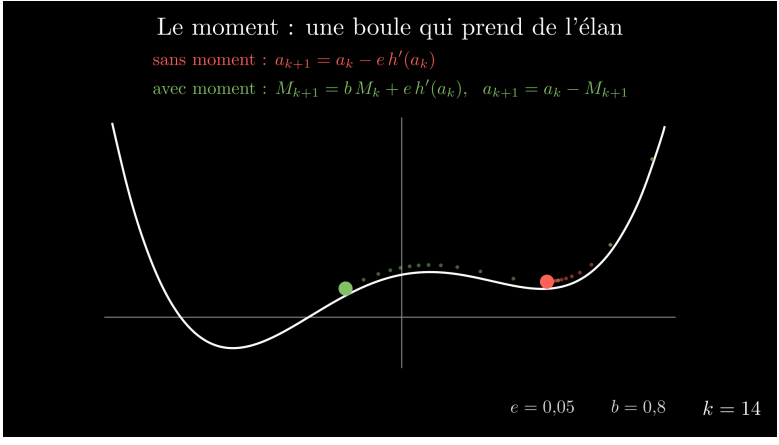
Le moment, en mouvement



Six itérations : la boule rouge descend dans la petite cuvette ; la verte, lancée, est déjà sur la bosse.

image 2 / 4 · ▶ la vidéo (21 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

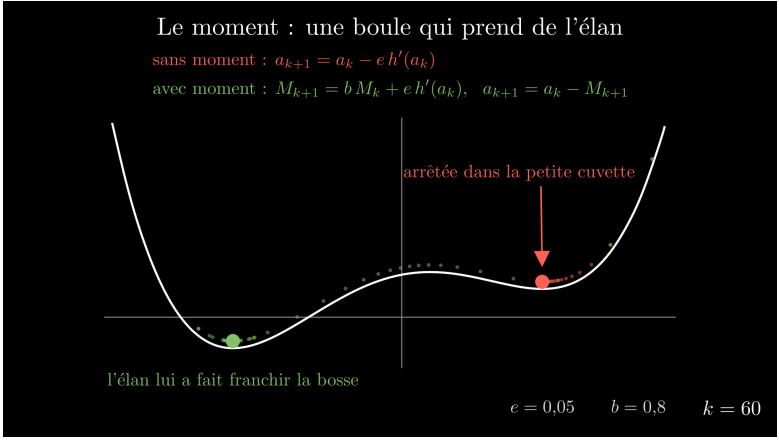
Le moment, en mouvement



Quatorze itérations : la verte a franchi la bosse ; la rouge ne bouge presque plus.

image 3 / 4 · ▶ la vidéo (21 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

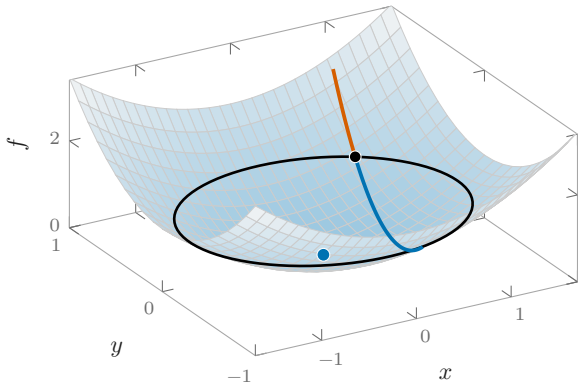
Le moment, en mouvement



La rouge reste dans la petite cuvette (un minimum local) ; la verte atteint la grande ($a \approx -1{,}08$).

image 4 / 4 · ▶ la vidéo (21 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

★ En deux variables : le gradient du TP 6



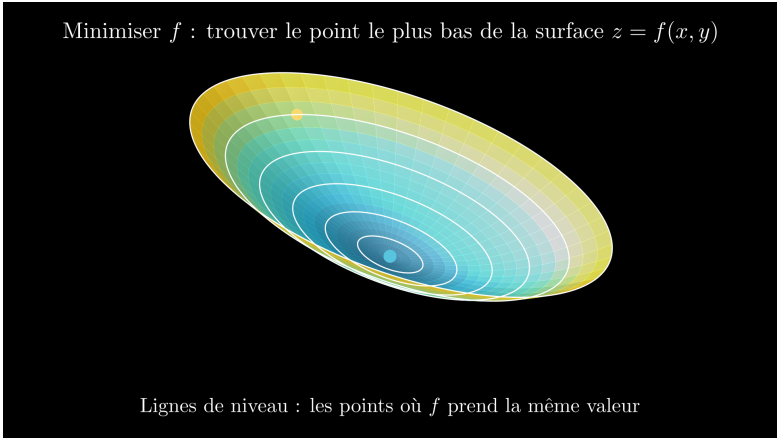
Pour $f(x, y)$, la pente est le **gradient** ∇f (TP 6), et la méthode devient

$$p_{k+1} = p_k - e\,\nabla f(p_k).$$

- Le long de la **ligne de niveau** (noir), f ne change pas.
- $-\nabla f(p)$ (bleu) : la direction où f descend le plus vite, perpendiculaire à la ligne de niveau. La courbe bleue descend, puis remonte : elle passe à côté du minimum.

Au TP (★) : la droite $y = a\,x + b$, à deux paramètres.

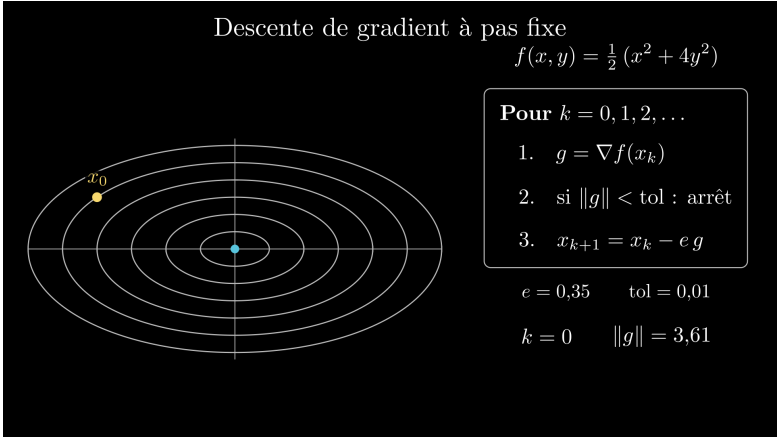
★ La méthode en deux variables, en mouvement



La surface $z = \frac{1}{2}(x^2 + 4y^2)$: une cuvette. Le départ x_0 (jaune), le minimum (bleu) au fond.

image 1 / 8 · ▶ la vidéo (53 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

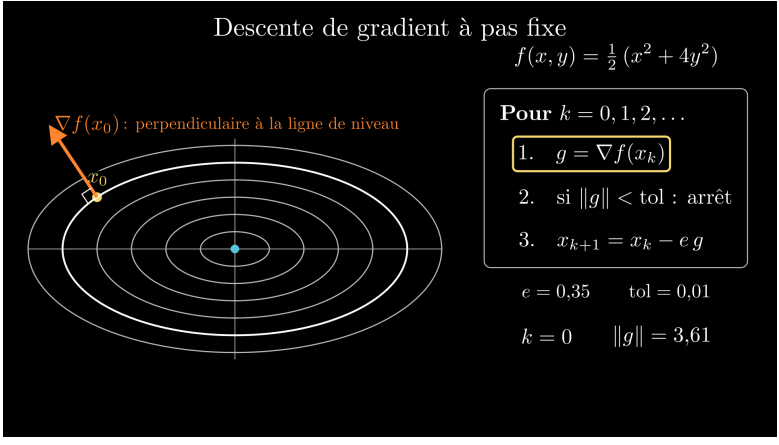
★ La méthode en deux variables, en mouvement



Vue de dessus : la carte des lignes de niveau. À droite, l'algorithme, avec un test d'arrêt.

image 2 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

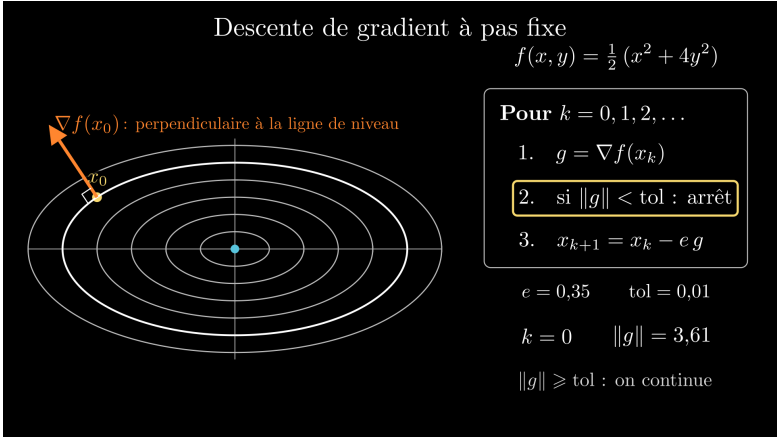
★ La méthode en deux variables, en mouvement



1. $g = \nabla f(x_0)$ (orange) : perpendiculaire à la ligne de niveau de x_0 (en blanc).

image 3 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

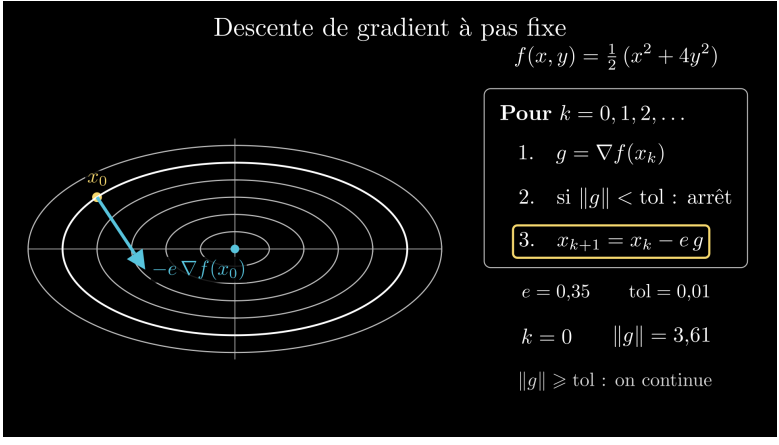
★ La méthode en deux variables, en mouvement



2. Test d'arrêt : $\|g\| \geqslant \text{tol}$, on continue.

image 4 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

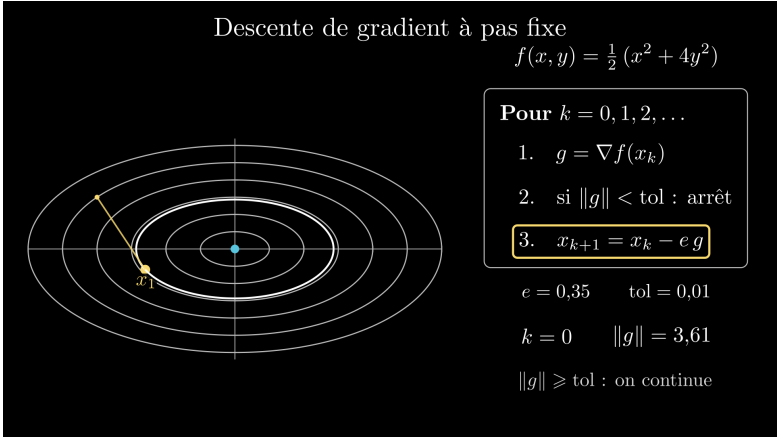
★ La méthode en deux variables, en mouvement



3. Le pas $-e\,g$ (bleu) : la direction de plus forte pente.

image 5 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

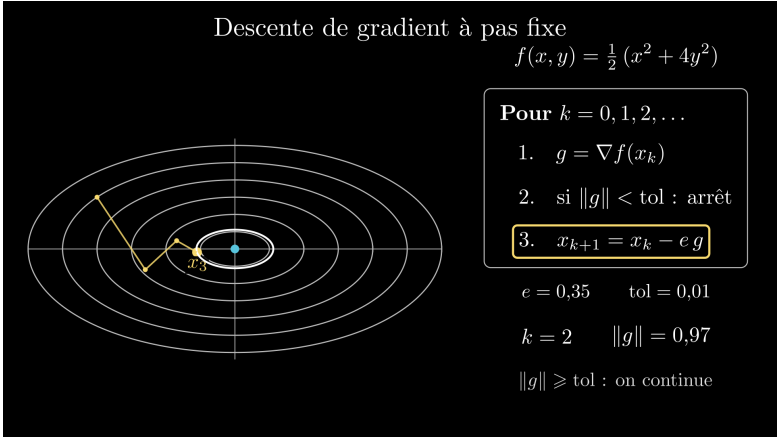
★ La méthode en deux variables, en mouvement



$x_1 = x_0 - e\,\nabla f(x_0)$, sur sa propre ligne de niveau ; on recommence en x_1.

image 6 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

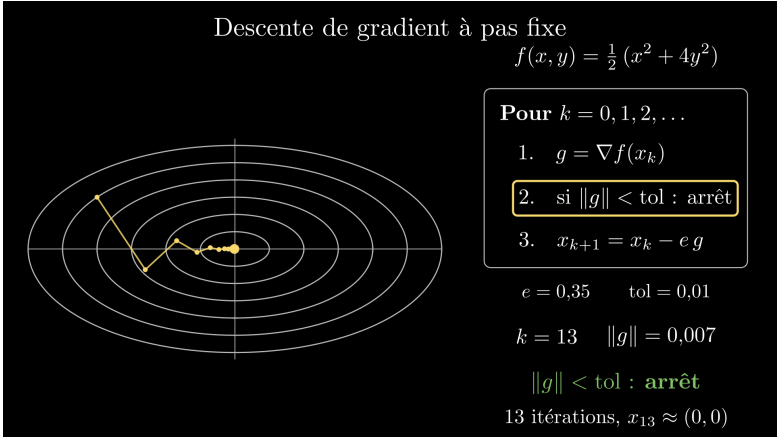
★ La méthode en deux variables, en mouvement



Après trois itérations : les pas raccourcissent, car $\|\nabla f(x_k)\|$ diminue près du minimum.

image 7 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

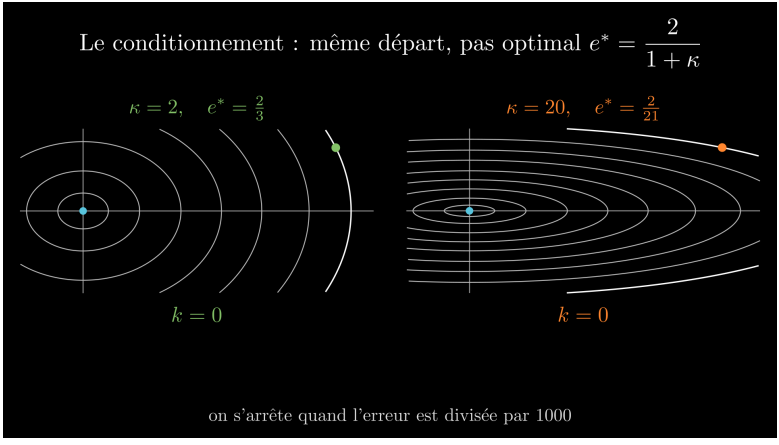
★ La méthode en deux variables, en mouvement



$\|g\| < \text{tol}$: arrêt. Treize itérations pour $e = 0{,}35$ et $\text{tol} = 10^{-2}$.

image 8 / 8 · ▶ la vidéo (53 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

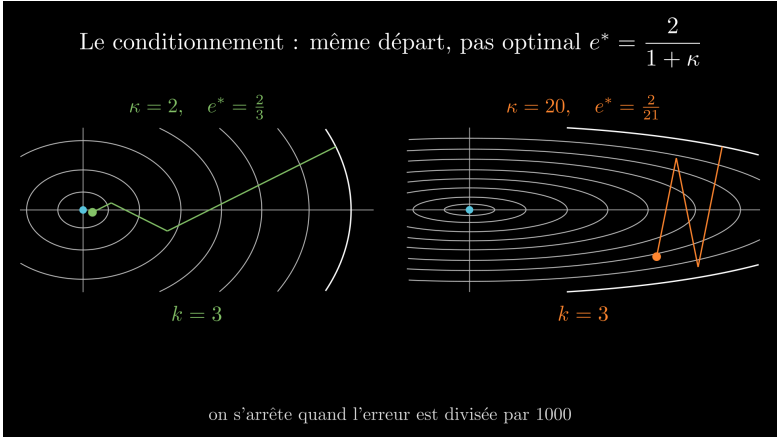
★ Une vallée étroite ralentit la méthode



$\frac{1}{2}(x^2 + \kappa y^2)$, même départ (sur la ligne de niveau blanche) et, pour chacune, le meilleur pas.

image 1 / 4 · ▶ la vidéo (22 s) : clic sur l'image (pdfpc), ou dans le lecteur vidéo

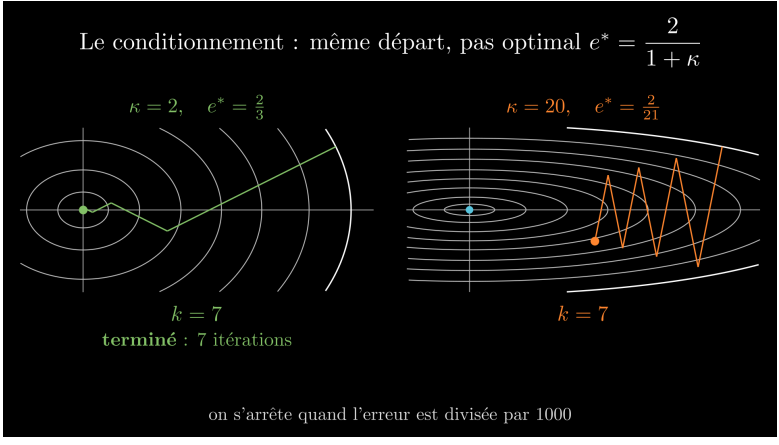
★ Une vallée étroite ralentit la méthode



Trois itérations : $\kappa = 2$ est presque arrivé ; $\kappa = 20$ zigzague d'un flanc à l'autre de la vallée.

image 2 / 4 · ▶ la vidéo (22 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

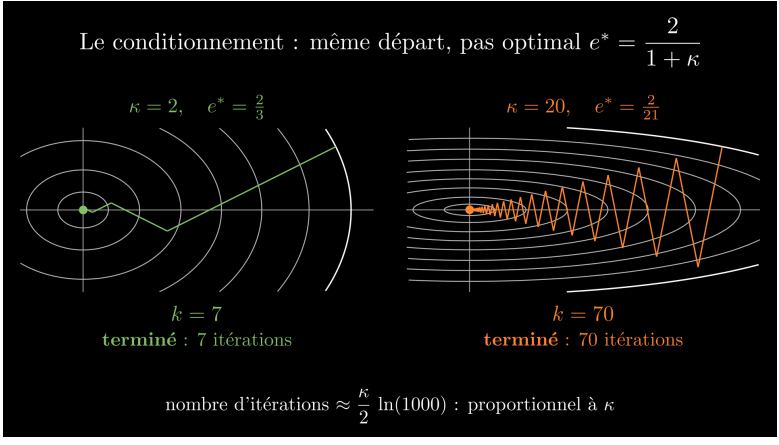
★ Une vallée étroite ralentit la méthode



$\kappa = 2$: terminé en 7 itérations.

image 3 / 4 · ▶ la vidéo (22 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

★ Une vallée étroite ralentit la méthode



$\kappa = 20$: 70 itérations. Au TP (★), des données non centrées donnent une telle vallée.

image 4 / 4 · ▶ la vidéo (22 s) : clic sur l'image 1 (pdfpc), ou dans le lecteur vidéo

Ce que vous allez programmer

```python
def gradient(f, a, e, n):
    points = [a]
    for k in range(n):
        a = ...    # avec derivee(f, a)
        points.append(a)
    return np.array(points)
```

Avant : `fun`, `pente`, `pentes_secantes`, `derivee`, `tangente` et `gradient_brutal`.

```python
def gradient_moment(f, a, e, b, n):
    points = [a]
    M = 0.0
    for k in range(n):
        M = ...
        a = ...
        points.append(a)
    return np.array(points)
```

Et la régression : `action`, `nuage_lineaire` ; ★ `grad_action`.

En résumé

1. La **dérivée** est la pente de la tangente, limite des pentes des sécantes. Numériquement : la différence centrée $\frac{f(a+h)-f(a-h)}{2h}$, d'erreur en h^2.
2. Pour descendre, on va dans le sens **opposé** au signe de $f'(a)$. Un pas fixe oscille autour du minimum ; un pas **proportionnel à la pente**, $a_{k+1} = a_k - e\,f'(a_k)$, converge.
3. **Le pas** : près d'un minimum, l'écart est multiplié par $1 - e\,f''(a^*)$; convergence si $0 < e < 2/f''(a^*)$. Trop petit : lent ; trop grand : oscillations.
4. La limite dépend du point de départ : c'est un minimum **local**.
5. **Moindres carrés** : la meilleure droite minimise la somme des carrés des résidus.
6. **Le moment** donne de l'élan et traverse les zones plates : c'est une boule pesante avec frottement, quand la méthode du gradient est la méthode d'Euler.

Pour aller plus loin : G. Goh, « Why Momentum Really Works », *Distill*, 2017 ; S. Boyd et L. Vandenberghe, *Introduction to Applied Linear Algebra*, chapitres 12 et 13 (moindres carrés), et *Convex Optimization*, § 9.2 et 9.3 (deux livres libres).

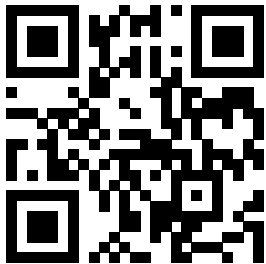
Au travail !

Au programme du TP 7

1. la dérivée : sécantes, tangente, dérivée numérique ;
2. le gradient brutal, puis la méthode du gradient, et le seuil du pas ;
3. la régression linéaire, sur un petit nuage puis sur des données fabriquées ;
4. le moment ;
5. ★ pour aller plus loin : la régression à deux paramètres.

Rappel

À l'ouverture, attendez le message ✓ **Boîte à outils chargée**. En fin de séance : **File** → **Download**.



storoo.fr/TP_EDO
dossier TP7, fichier TP7_DescenteGradient.ipynb