Ift 2421

Chapitre 1

Chiffres significatifs
et
propagation d’erreurs

Sources d'erreurs

Erreurs de données

1. Erreurs de modèles

Exemple:

- Ignorer la viscosité dans un modèle en mécanique des fluides.
- Utiliser la mécanique classique plutôt que quantique en physique atomique.
- Loi de Hooke (modèle linéaire) pour un ressort $F = k\,x$ alors que $F = k(x)\,x$ (modèle non linéaire).

2. Erreurs de mesures

Exemple:

- la précision de ce thermomètre est de ½ degré Celcius .
- ce voltmètre offre une précision de 0.0001 volt.

Erreurs numériques

1. Erreurs de troncature

Exemple:

- $$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \approx \sum_{n=0}^{100} \frac{x^n}{n!}$$

- Un processus infini est tronqué afin d'obtenir une procédure finie => Erreurs

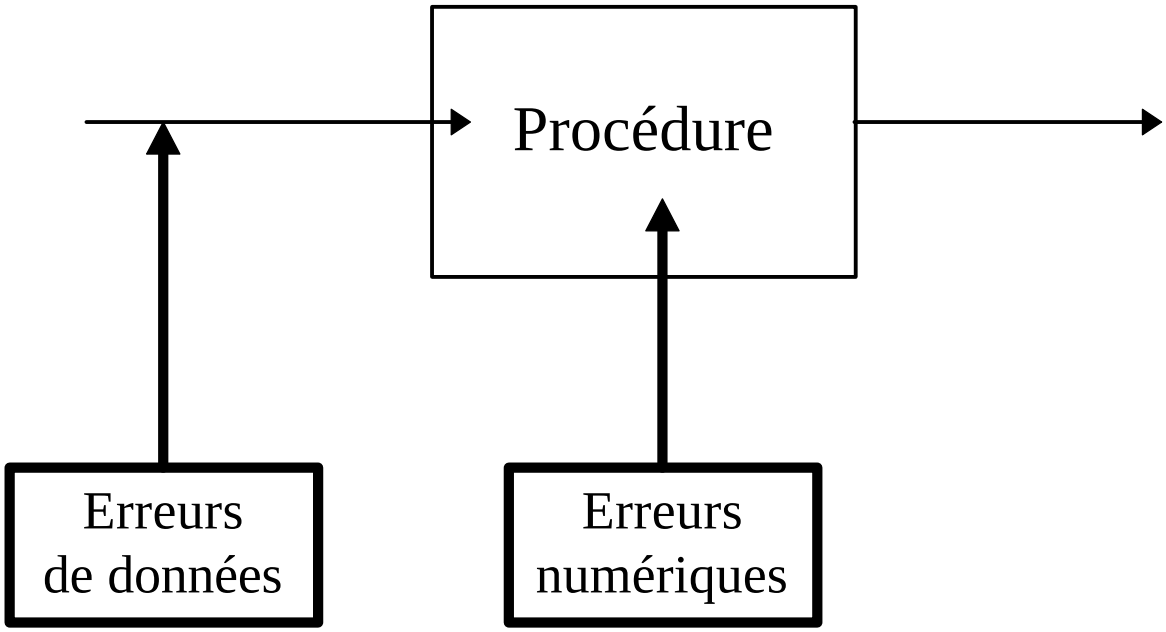
2. Erreurs d'arrondis (d'affectation)

Que vaut 0.1 en base 2?

$$(0.1)_{10} = (0.0001100110011\ldots)_2$$

Cela ne peux pas être conservé exactement dans l'ordinateur!

$\Re$ ne peux pas être représenté dans l'ordinateur.



Arithmétique des ordinateurs

Notation flottante:

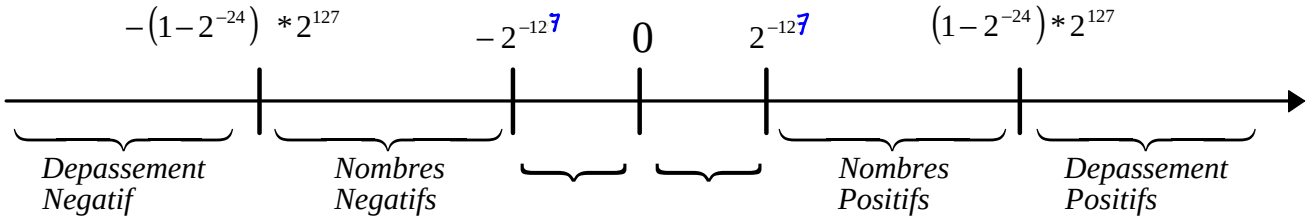
$$fl(X) = \pm .d_1 d_2 d_3 \ldots d_s * b^e$$

b est la base
e est l'exposant avec $m \leq e \leq M$
s est le nombre de chiffres de la mantisse
et où

$$0 \leq d_i \leq b-1 \; et \; d_1 \neq 0$$

Ceci nous assure l'unicité de la représentation.

Un système de numérotation flottante est défini par le quadruplet
(b,s,m,M)



Représentation des nombres dans un ordinateur

Le nombre de bits pour représenter les nombres en point flottant est fini.

Le nombre de valeurs distinctes représentables dans un ordinateur est donc lui aussi fini.

Exemple:

Soit le système de numérotation flottante défini par:

$b = 2$, $s = 2$ et $-2 \leq e \leq 3$

Dans ce système, les nombres normalisés sont de la forme

$\pm .10_2 * 2^e$ ou $\pm .11_2 * 2^e$ avec $-2 \leq e \leq 3$

Voici la liste de tous les nombres positifs représentables dans ce système.

$$.10_2 * 2^{-2} = \frac{1}{8} \qquad .11_2 * 2^{-2} = \frac{3}{16}$$

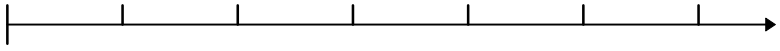
$$.10_2 * 2^{-1} = \frac{1}{4} \qquad .11_2 * 2^{-1} = \frac{3}{8}$$

$$.10_2 * 2^{0} = \frac{1}{2} \qquad .11_2 * 2^{0} = \frac{3}{4}$$

$$.10_2 * 2^{1} = 1 \qquad .11_2 * 2^{1} = \frac{3}{2}$$

$$.10_2 * 2^{2} = 2 \qquad .11_2 * 2^{2} = 3$$

$$.10_2 * 2^{3} = 4 \qquad .11_2 * 2^{3} = 6$$



Erreurs de calcul en points flottant:

Représentation:

Exemple:
code Mathematica

```
x=1.0;
While[ x != 0.0,
       Print[x];
       x = x -0.1
     ]
```

Précision:

Epsilon machine:

Pour chaque machine, de la plus petite calculatrice au plus grand ordinateur, on peut associer un nombre positif,

EPS = ″Epsilon Machine″

Par définition, Eps est le plus petit nombre positif tel que:

$$1 + EPS \neq 1$$

ou encore

$$(1 + EPS) - 1 \neq 0$$

Troncature:

Exemple avec b=10, s=3

$$X = 1237.$$

$$fl(X) = fl(+ 0.1237 * 10^4) = + 0.123 * 10^4$$

La troncature est toujours biaisée puisque

$$fl(X) \leq X$$

Arrondissement:

Même exemple.

$$X = 1237.$$

$$fl(X) = fl(+ 0.1237 * 10^4) = + 0.124 * 10^4$$

L'arrondissement est non biaisé et, tour à tour, on

$$fl(X) \leq X \qquad \text{(50\% du temps)}$$

et

$$fl(X) \geq X \qquad \text{(50\% du temps)}$$

Procédure pour tronquer à s chiffres

$$X = \pm\ .d_1 d_2 d_3 \ldots d_s d_{s+1} \ldots * b^e$$

$$tron(X) = \pm\ .d_1 d_2 d_3 \ldots d_s\ * b^e$$

Procédure pour arrondir à s chiffres

$$X = \pm\ .d_1 d_2 d_3 \ldots d_s d_{s+1} \ldots * b^e$$

$$X^* = \pm\ .d_1 d_2 d_3 \ldots d_s d_{s+1} \ldots * b^e$$

$$\pm\ .0\ \ 0\ \ 0 \ldots 0 v_{s+1} \ldots * b^e$$

$$v_{s+1} = b/2$$

Arrondi(X) = tron(X*)

Propriétés des 4 opérations

X+Y → fl(fl(X) + fl(Y))

X-Y → fl(fl(X) - fl(Y))

X*Y → fl(fl(X) * fl(Y))

X/Y → fl(fl(X) / fl(Y))

On arrondit ou on tronque?

Pour les exemples, b=10, s=3.

Multiplication:

$(1/3) \times 3 \rightarrow$

$fl(fl(1/3) \times fl(3)) =$
$fl((.333 * 10^0) \times (0.300 * 10^1)) =$
$fl(0.0999 * 10^1) =$
$0.999 * 10^0$

Si l'exposant n'est pas le même, on peut perdre de la précision à cause de l'erreur de **décalage**.

Addition

$1.37 + 0.0269 \rightarrow .137 * 10^{1} + .269 * 10^{-1}$

$$\begin{array}{rl} .137 & *10^{1} \\ + \ .00269 & *10^{1} \\ \hline .13969 & *10^{1} \end{array}$$

Tronqué → $.139 * 10^{1}$
Arrondi → $.140 * 10^{1}$

Soustraction

$4850 - 4820 \rightarrow .485 * 10^{4} - .482 * 10^{4}$

$$\begin{array}{rl} .485 & *10^{4} \\ - \ .482 & *10^{4} \\ \hline .003 & *10^{4} \\ .300 & *10^{2} \end{array}$$

Tronqué → $.300 * 10^{2}$
Arrondi → $.300 * 10^{2}$

Soustraction (Autre exemple)

$3780 - .321 \rightarrow .378 * 10^4 - .321 * 10^0$

$$\begin{array}{rl} .378 & *10^4 \\ -\ .0000321 & *10^4 \\ \hline .3779679 & *10^4 \end{array}$$

Tronqué → $.377 * 10^4$

Arrondi → $.378 * 10^4$

Multiplication (Autre exemple)

$403000 * .0197 \rightarrow .403 * 10^6 * .197 * 10^{-1}$

$$\begin{array}{rr} .403 & 6 \\ * \ .197 & +\ -1 \\ \hline .079391 & 5 \end{array}$$

$.079391 * 10^5$

$.79391 * 10^4$

Tronqué → $.793 * 10^4$

Arrondi → $.794 * 10^4$

Attention:

Addition

$1.3685 + 0.0269 \rightarrow$

$fl(fl(.13685 * 10^{1)} + fl(.269 * 10^{-1}))=$

Tronqué $\rightarrow .138 * 10^{1}$

Arrondi $\rightarrow .140 * 10^{1}$

Pour une suite d'opération, il faut arrondir ou tronqué chaque résultat intermédiaire.

L'ordre dans lequel sont effectuées les opérations est très important.

On n'a plus l'associativité des opérations + et *.

Erreurs sur les données:

Définition:

Q valeur exacte
($\pi = 3.14159265...$)

Q^* approximation de Q
($\pi^* = 3.14$)

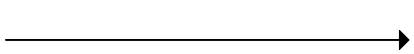
Comment définir l'erreur sur Q?

Erreur absolue

$$\Delta Q = \left| Q - Q^* \right|$$

Intervalle de confiance:

l'intervalle de largeur $2\Delta Q$ et de centre Q^*.



Erreur relative

$$\Delta_r(Q) = \frac{\left| Q - Q^* \right|}{\left| Q \right|}$$

On emploie plutôt

$$E_r(Q) = \frac{\left| Q - Q^* \right|}{\left| Q^* \right|}$$

Chiffres significatifs exacts (cse)

Un chiffre significatif d'une valeur Q^* est exact
si l'erreur absolue (ΔQ)
sur cette valeur est

$\leq$ ½ fois l'unité du rang du chiffre.

Exemple:

$$Q = 3.2189 \pm 0.0003$$

Le '8' est- il un cse?

Rang du 8 = -3
Unité du rang du 8 = 0.001
½ fois cette unité = 0.0005

$$\Delta Q = 0.0003 \leq 0.0005$$

le 8 est un cse et c'est le dernier.
Q a donc 4 cse.

$$Q = \underline{3.2189} \pm 0.0003$$

(souvent on souligne les cse)

Conséquences:

La $n^{ième}$ décimale d'une valeur est exacte

$\Leftrightarrow$

$\Delta Q \leq 0.5 \times 10^{-n}$

Le $n^{ième}$ chiffre devant le point est exact

$\Leftrightarrow$

$\Delta Q \leq 0.5 \times 10^{n-1}$

Remarques:

- Si tous les chiffres sont exacts, une borne pour l'erreur absolue est égale à ½ fois l'unité du dernier chiffre.
- Souvent aussi, lorsque l'erreur absolue n'est pas connue, on suppose que le dernier chiffre n'est pas exact.

Exemple:
Combien y a t il de cse dans la valeur approchée de π donnée par

$$\pi^* = 22 / 7$$

$$\pi^* = 3.1428571...$$

$$\Delta Q = | \pi - \pi^* | = 0.0012644... \leq 0.00\ 5 \leq 0.5\ 10^{-2}$$

Le rang du dernier cse est -2.

$$\pi^* = \underline{3.14}28571...$$

on écrit en général $\underline{3.14}3 \pm 0.005$

Combien y a t il de cse dans la valeur approchée de π donnée par

$$\pi^* = 3.1416$$

$$\Delta Q = | \pi - \pi^* | = 0.73\ 10^{-5}$$

$$\pi^* = 3.1416$$

Cancellation de ‘cse’

Soustraction de valeurs presque égales

$$U = \sqrt{7001} \quad U^* = 0.83672 \times 10^2 \ (avec 5 cse)$$
$$V = \sqrt{7000} \quad V^* = 0.83666 \times 10^2 \ (avec 5 cse)$$

$$X = U - V = 0.59759 \times 10^{-2}$$
$$X^* = U^* - V^* = 0.6 \times 10^{-2}$$

$$\Delta X = 0.24 \times 10^{-4}$$
$$\Rightarrow X = 0.6 \times 10^{-2} \pm 0.0024 \times 10^{-2}$$

Plus que 4 cse, problème

Remarque:

$$X = \sqrt{a} - \sqrt{b}$$

or

$$(\sqrt{a} - \sqrt{b}) \times (\sqrt{a} + \sqrt{b}) = (a - b)$$

donc

$$X = \frac{a - b}{\sqrt{a} + \sqrt{b}}$$

et

$$X^* = \frac{1}{1.6734 \times 10^2} = 0.59759 \times 10^{-2}$$

5cse!

Attention

- à l’ordre d’évaluation des opérations.
- aux simplification à apporter pour minimiser les erreurs.

Séries de Taylor

Fonction à une variable

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R(x,a)$$

$$R(x,a) = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)dt = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(x\ )$$

Fonction à plusieurs variables

$$f(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) + \frac{f_{xx}(a,b)(x-a)^2 + f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2}{2!} + \ldots$$

Propagation d'erreurs

$$X = X^* \pm \Delta X$$

$$F(X) = F(X^* \pm \Delta X) = F(X^*) \pm \Delta F$$

Taylor au premier ordre

$$F(X) \approx F(X^*) + F'(X^*)(X - X^*) + \ldots$$

$$\Rightarrow F(X) \approx F(X^*) \pm \left|F'(X^*)\right| \ \Delta X$$

donc $\Delta F = \left|F'(X^*)\right| \ \Delta X$ (Variable au premier ordre)

Exemple

Si F(X) = $X^{1/2}$ et X^* = <u>1.234</u> (avec 4 cse)

alors X = 1.234 ± 0.0005

et F(X^*) = 1.11086

$$F'(X) = \left(\frac{1}{2} X^{-\frac{1}{2}}\right)$$

$$\Delta F = \left|F'(X^*)\right| \ \Delta X = (0.0005)(0.4501) = 0.0002251$$

F(X) = 1.1109 ± 0.0002251

4 cse aussi

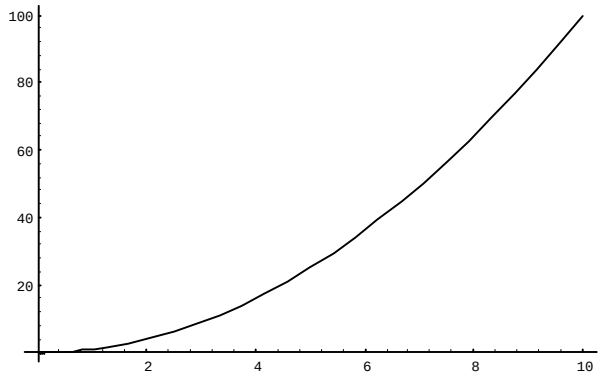
Méthode de la fourchette

Si $X \in [a, b]$

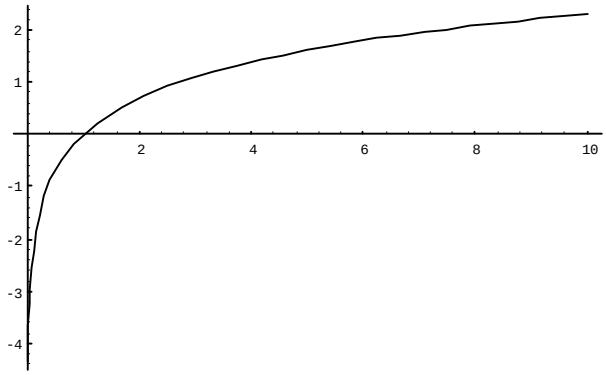
et si f est continue monotone sur [a, b]

alors

$F[X] \in [\ F(a), F(b)\ ]$



$F(X) = X^2$



$F(X) = \ln(X)$

Exemple

Si $F(X) = X^{1/2}$ et $X^* = \underline{1.234}$ (avec 4 cse)

alors $X = 1.234 \pm 0.0005$

$X \in [1.2335, 1.2345]$

$F(1.2335) = 1.11063$
$F(1.2345) = 1.11108$

donc $F(X) \in [1.11063, 1.11108]$

$F(X) = 1.11086 \pm 0.000225052$

$F(X) = 1.1109 \pm 0.0003$
4 cse aussi

Propagation de l'erreur
dans les fonctions a plusieurs variables

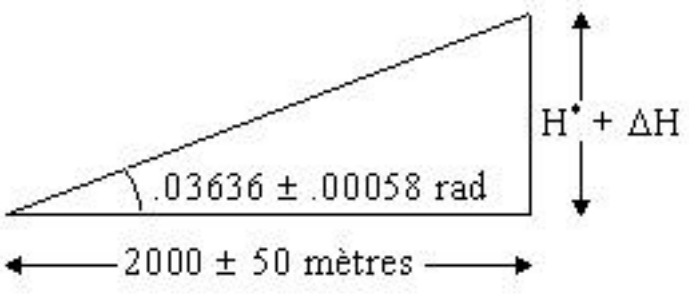
$$F(X,Y) \simeq F(X^*,Y^*) + F_X(X^*,Y^*) \times (X - X^*) + F_Y(X^*,Y^*) \times (Y - Y^*)$$

l'intervalle de confiance de F(X,Y) est alors donné par

$$F(X,Y) = F(X^*,Y^*) \pm \Delta F$$

avec ΔF approché au 1er ordre par

$$\Delta F = \left|F_X(X^*,Y^*)\right| \Delta X + \left|F_Y(X^*,Y^*)\right| \Delta Y$$



Trouver $H^* + \Delta H$

Ici $H^* = L^* \tan(\theta^*)$

avec $L^* = 2000$ m
et $\theta^* = 0.03636$ radians

donc $H^* = 2000 \tan(0.03636) = 72.8$

$$\Delta H \approx \left|H_L(L^*, q^*)\right|\Delta L + \left|H_q(L^*, q^*)\right|\Delta q$$

$$= \left|\tan(q^*)\right|\Delta L + \left|\frac{L^*}{\cos^2(q^*)}\right|\Delta q$$

$$= \left|\tan(0.03636)\right| \times 50 + \left|\frac{2000}{\cos^2(0.03636)}\right| \times 0.00058$$

$$\approx 3.0$$

donc $H = H^* \pm \Delta H = 73 \pm 3$

1 seul cse

Addition

$$F^* = F(U^*, V^*) = U^* + V^*$$

$$\Delta F \approx |1|\Delta U + |1|\Delta V$$
$$= \Delta U + \Delta V$$

Soustraction

$$F^* = F(U^*, V^*) = U^* - V^*$$

$$\Delta F \approx |1|\Delta U + |-1|\Delta V$$
$$= \Delta U + \Delta V$$

Multiplication

$$F^* = F(U^*, V^*) = U^* \times V^*$$

$$\Delta F \approx |V^*|\Delta U + |U^*|\Delta V$$

$$\frac{\Delta F}{|F^*|} = \frac{\Delta U}{|U^*|} + \frac{\Delta V}{|V^*|}$$

Division

$$F^* = F(U^*, V^*) = U^* \div V^*$$

$$\Delta F \approx \frac{\Delta U}{|V^*|} + \left|\frac{-U^*}{(V^*)^2}\right|\Delta V$$

$$\frac{\Delta F}{|F^*|} = \frac{\Delta U}{|U^*|} + \frac{\Delta V}{|V^*|}$$

L'erreur absolue sur la somme ou la différence de deux valeurs est la somme des erreurs absolues.

L'erreur relative sur la multiplication ou la division de deux valeurs est la somme des erreurs relatives.

Calculer $F(X,Y) = X^2 - Y^2$

avec

$X = 10 \pm 0.05$ et $Y = 2 \pm 0.03$

$F^* = F(X^*,Y^*) = 96$

$\Delta F = |2\ X|\ \Delta X + |2\ Y|\ \Delta Y$

$= (20)\ (0.05) + (4)\ (0.03) = 1.12$

donc $F = 96 \pm 1.12$

or

$F(X,Y) = (X+Y)\ (X-Y) = U\ V$

$U = X+Y$ donc $\Delta U = \Delta X + \Delta Y$

$V = X-Y$ donc $\Delta V = \Delta X + \Delta Y$

et

$\Delta F = |V|\ \Delta U + |U|\ \Delta V$

$= (8)\ (0.08) + (12)\ (0.08) = 1.6$

donc $F = 96 \pm 1.6$

Annexe au chapitre 1 :

Arrondissage et CSE :

Cas	Arrondi	Erreur d'arrondi
1	$X \to X^*$	$\left\|X - X^*\right\|$
2	$X_1^* \to X_2^*$	$\left\|X_1^* - X_2^*\right\|$

1. $X \to X^*$

Si un nombre exact X est arrondi à n chiffres alors on obtient un nombre approché X^* de n chiffres significatifs exacts.

$$X = x_1\ x_2\ x_3\ \ldots\ x_i\ \ldots\ x_{n-2}\ x_{n-1}\ \mathbf{x_n}\ x_{n+1}\ x_{n+2}\ \ldots$$

$$X^* = x_1\ x_2\ x_3\ \ldots\ x_i\ \ldots\ x_{n-2}\ x_{n-1}\ \mathbf{x_n}\ 0\ 0\ 0\ 0\ \ldots$$

Soit k le rang de x_{n+1}.

Arrondissage à n chiffres $\Rightarrow \left|X - X^*\right| \leq 5\,10^k = 0.5\,10^{k+1}$

Donc le rang du dernier CSE est k+1 : c'est le rang de x_n.

Il y a donc n CSE dans X^*.

2. $X_1^* \rightarrow X_2^*$

Si un nombre approché X_1^* de n chiffres significatifs exacts est arrondi à n chiffres alors le nouveau nombre approché X_2^* possède n-1 chiffres significatifs exacts garantis.

Soit k le rang de x_{n+1}.

X_1^* est une valeur approchée de X.

X_1^* à n CSE donc $\left|X - X_1^*\right| \leq 5\,10^k$

X vraie valeur

X_1^* = $x_1\ x_2\ x_3 \ldots x_i \ldots x_{n-2}\ x_{n-1}\ \mathbf{x_n}\ x_{n+1}\ x_{n+2} \ldots$

X_2^* = $x_1\ x_2\ x_3 \ldots x_i \ldots x_{n-2}\ x_{n-1}\ \mathbf{x_n}\ 0\ 0\ 0\ 0 \ldots$

X_1^* est arrondi à n chiffres donc $\left|X_1^* - X_2^*\right| \leq 5\,10^k$

La question est donc $\left|X - X_2^*\right| \leq$???

$$\begin{aligned}\left|X - X_2^*\right| &= \left|X - X_1^* + X_1^* - X_2^*\right| \\ &\leq \left|X - X_1^*\right| + \left|X_1^* - X_2^*\right| \\ &\leq 5\,10^k + 5\,10^k \\ &\leq 10\,10^k = 10^{k+1} \\ &\leq .5\,10^{k+2}\end{aligned}$$

Donc le rang du dernier CSE est k+2 : c'est le rang de x_{n-1}.

Il y a donc n-1 CSE garantis dans X_2^*.