Tutorial 3

Exercice 1

1) Montrons que $\text{cond}(\Sigma \rho_i) = \dfrac{\sum_{i=1}^{n} |\rho_i|}{\left|\sum_{i=1}^{n} \rho_i\right|}$.

$$\left|\sum_{i=1}^{n} \tilde{\rho}_i - \sum_{i=1}^{n} \rho_i\right| = \left|\sum_{i=1}^{n} (\tilde{\rho}_i - \rho_i)\right| \leq \sum_{i=1}^{n} |\tilde{\rho}_i - \rho_i|$$
$$\leq \sum_{i=1}^{n} \varepsilon |\rho_i| \leq \varepsilon \sum_{i=1}^{n} |\rho_i|$$

On en déduit que :

$$\frac{\left|\sum_{i=1}^{n} \tilde{\rho}_i - \sum_{i=1}^{n} \rho_i\right|}{\varepsilon \left|\sum_{i=1}^{n} \rho_i\right|} \leq \frac{\varepsilon \sum_{i=1}^{n} |\rho_i|}{\varepsilon \left|\sum_{i=1}^{n} \rho_i\right|} = \frac{\sum_{i=1}^{n} |\rho_i|}{\left|\sum_{i=1}^{n} \rho_i\right|}$$

On a donc $\text{cond}(\Sigma \rho_i) \leq \dfrac{\sum_{i=1}^{n} |\rho_i|}{\left|\sum_{i=1}^{n} \rho_i\right|}$

Si on choisit $\tilde{\rho}_i = \rho_i + \varepsilon \, \text{sign}(\rho_i) \, \rho_i$

on a $|\tilde{\rho}_i - \rho_i| = |\varepsilon \, \text{sign}(\rho_i) \rho_i| = \varepsilon |\rho_i|$

$$\frac{\left|\sum_{i=1}^{n}\tilde{\rho}_i - \sum_{i=1}^{n}\rho_i\right|}{\varepsilon\left|\sum_{i=1}^{n}\rho_i\right|} = \frac{\left|\sum_{i=1}^{n}\varepsilon\,\mathrm{sign}(\rho_i)\rho_i\right|}{\varepsilon\left|\sum_{i=1}^{n}\rho_i\right|}$$

$$= \frac{\left|\sum_{i=1}^{n}\mathrm{sign}(\rho_i)\rho_i\right|}{\left|\sum_{i=1}^{n}\rho_i\right|} = \frac{\sum_{i=1}^{n}|\rho_i|}{\left|\sum_{i=1}^{n}\rho_i\right|}$$

En conclusion, $\mathrm{cond}\left(\sum \rho_i\right) = \dfrac{\sum_{i=1}^{n}|\rho_i|}{\left|\sum_{i=1}^{n}\rho_i\right|}$.

2) Algorithme de sommation récursif

```
function res = Sum(ρ)
    σ_0 = 0
    for i = 1 : n
        σ_i = σ_{i-1} ⊕ ρ_i
    end
    res = σ_n
```

Modèle standard : $fl(x+y) = x \oplus y = (1+\delta)(x+y)$.

avec $|\delta| \leq u$

$\sigma_0 = 0$

$\sigma_1 = p_1$

$\sigma_2 = \sigma_1 \oplus p_2 = p_1 \oplus p_2 = (1+\delta_1)(p_1+p_2)$
avec $|\delta_1| \leq u$

$\sigma_3 = \sigma_2 \oplus p_3 = [\sigma_2 + p_3](1+\delta_2)$ avec $|\delta_2| \leq u$.

$= [(p_1+p_2)(1+\delta_1) + p_3](1+\delta_2)$

$= (p_1+p_2)(1+\delta_1)(1+\delta_2) + p_3(1+\delta_2)$

$\sigma_4 = \sigma_3 \oplus p_4 = [\sigma_3 + p_4](1+\delta_3)$ avec $|\delta_3| \leq u$

$= [(p_1+p_2)(1+\delta_1)(1+\delta_2) + p_3(1+\delta_2) + p_4]$
$\times (1+\delta_3)$

$= (p_1+p_2)(1+\delta_1)(1+\delta_2)(1+\delta_3) + p_3(1+\delta_2)(1+\delta_3)$
$+ p_4(1+\delta_3)$

⋮ ⋮ ⋮

$$\sigma_n = (p_1+p_2)\prod_{i=1}^{n-1}(1+\delta_i) + p_3\prod_{i=2}^{n-1}(1+\delta_i) + p_4\prod_{i=3}^{n-1}(1+\delta_i)$$
$$+ \dots + p_n(1+\delta_{n-1}).$$

On peut montrer que, $|\delta_i| \le u$

$$\prod_{i=1}^{n}(1+\delta_i) = 1+\theta_n \quad \text{avec} \quad |\theta_n| \le \gamma_n = \frac{nu}{1-nu}$$

$$\sigma_n = (p_1+p_2)(1+\theta_{n-1}) + p_3(1+\theta_{n-2}) + p_4(1+\theta_{n-3})$$
$$+ p_n(1+\theta_1).$$

On en déduit que σ_n est la somme exacte du vecteur $\tilde{p} = [\underbrace{p_1(1+\theta_{n-1})}_{\tilde{p}_1}, \underbrace{p_2(1+\theta_{n-1})}_{\tilde{p}_2}, \underbrace{p_3(1+\theta_{n-2})}_{\tilde{p}_3}$
$\dots, \underbrace{p_n(1+\theta_2)}_{\tilde{p}_n}]$

$$|\tilde{p}_i - p_i| = |p_i(1+\theta_{n-i+1}) - p_i|$$
$$\le |\theta_{n-i+1}|\,|p_i| \le \gamma_{n-1}|p_i|$$

L'erreur inverse :

$$\eta(\sigma_n) = \max_{i=1:n} \frac{|\tilde{p}_i - p_i|}{p_i} \le \gamma_{n-1} \approx (n-1)u$$

L'algorithme est bien **inverse-stable**.

3) On cherche une borne pour $|\sigma_n - \sum_{i=1}^{n} \rho_i|$

$$\left| (\rho_1+\rho_2)(1+\theta_{n-1}) + \rho_3(1+\theta_{n-2}) + \rho_4(1+\theta_{n-3}) + \rho_n(1+\theta_1) - \sum_{i=1}^{n} \rho_i \right| \quad (\sigma_n)$$

$$\leq \left| (\rho_1+\rho_2)\theta_{n-1} + \rho_3\theta_{n-2} + \dots + \rho_n\theta_1 \right|$$

$$\leq |\theta_{n-1}| \sum_{i=1}^{n} |\rho_i| \leq \gamma_{n-1} \sum_{i=1}^{n} |\rho_i|.$$

Donc $\left|\sigma_n - \sum_{i=1}^{n} \rho_i\right| \leq \gamma_{n-1} \sum_{i=1}^{n} |\rho_i|$

En divisant par $|\Sigma \rho_i|$, on obtient

$$\frac{\left|\sigma_n - \sum_{i=1}^{n} \rho_i\right|}{\left|\sum_{i=1}^{n} \rho_i\right|} \leq \gamma_{n-1} \frac{\sum_{i=1}^{n} |\rho_i|}{\left|\sum_{i=1}^{n} \rho_i\right|}$$

$$\leq \underbrace{\gamma_{n-1}}_{\approx (n-1)u} \operatorname{cond}(\Sigma \rho_i)$$

4) Mêmes questions avec le produit scalaire.

$x = (x_i)$, $y = (y_i) \in \mathbb{R}^n$.

$$x^T y = \sum_{i=1}^{n} x_i y_i$$

* Montrer que $\text{cond}(x^T y) = \dfrac{2|x|^T|y|}{|x^T y|}$.

Par définition

$$\text{cond}(x^T y) = \lim_{\varepsilon \to 0} \sup \left| \frac{\tilde{x}^T \tilde{y} - x^T y}{\varepsilon\, x^T y} \right| :$$

$$\left.\begin{aligned} |\tilde{x}_i - x_i| &\le \varepsilon |x_i| \\ |\tilde{y}_i - y_i| &\le \varepsilon |y_i| \end{aligned}\right\} \quad i = 1 \dots n$$

$$\frac{|\tilde{x}^T \tilde{y} - x^T y|}{\varepsilon\, |x^T y|} = \left| \frac{(x+\Delta x)^T (y+\Delta y) - x^T y}{\varepsilon\, x^T y} \right|$$

$$= \left| \frac{x^T y + (\Delta x)^T y + x^T \Delta y + (\Delta x)^T \Delta y - x^T y}{\varepsilon\, x^T y} \right|$$

$$\le \frac{1}{\varepsilon} \left(\frac{|(\Delta x)^T y| + |x^T \Delta y| + |(\Delta x)^T \Delta y|}{|x^T y|} \right)$$

$$\le \frac{1}{\varepsilon} \left(\frac{|\Delta x|^T |y| + |x|^T |\Delta y| + |\Delta x|^T |\Delta y|}{|x^T y|} \right)$$

$$\le \frac{1}{\varepsilon} \left(\frac{\varepsilon |x|^T |y| + \varepsilon |x|^T |y| + \varepsilon^2 |x|^T |y|}{|x^T y|} \right)$$

$$\le \frac{2|x|^T |y| + \varepsilon |x|^T |y|}{|x^T y|}$$

$$\le 2 \frac{|x|^T |y|}{|x^T y|} \quad (1)$$

$$\leq \frac{|\Delta x|^T |y| + |x|^T |\Delta y| + |\Delta x|^T \Delta y}{\varepsilon |x^T y|}.$$

Or $|\Delta x| \leq \varepsilon |x|$ et $|\Delta y| \leq \varepsilon |y|$

D'où

① $$\leq \frac{\varepsilon |x|^T |y| + \varepsilon |x|^T |y| + \varepsilon^2 |x|^T |y|}{\varepsilon |x^T y|}$$

$$\leq \frac{2|x|^T |y|}{|x^T y|} + \varepsilon \frac{|x|^T |y|}{|x^T y|}.$$

$$\Rightarrow \text{cond}(x^T y) \leq \frac{2|x|^T |y|}{|x^T y|}.$$

Prenons

$$\tilde{x}_i = x_i + \varepsilon \, \text{sign}(y_i) |x_i|$$
$$\tilde{y}_i = y_i + \varepsilon \, \text{sign}(x_i) |y_i|$$

$$\tilde{x}_i \tilde{y}_i = x_i y_i + \varepsilon x_i \, \text{sign}(x_i) |y_i| + \varepsilon y_i \, \text{sign}(y_i) |x_i| + \varepsilon^2 \, \text{sign}(y_i) \, \text{sign}(x_i) |x_i| |y_i|$$

$$= x_i y_i + 2\varepsilon |x_i| |y_i| + \varepsilon^2 \, \text{sign}(x_i y_i) |x_i| |y_i|$$

$$\left|\frac{\tilde{x}^T\tilde{y} - x^Ty}{\varepsilon\, x^Ty}\right| = \frac{2\varepsilon|x^T||y| + \varepsilon^2\sum_{i=1}^{n}\operatorname{sign}(x_iy_i)|x_i||y_i|}{|\varepsilon\, x^Ty|}$$

En passant à la limite quand $\varepsilon \to 0$

on obtient $\operatorname{cond}(x^Ty) = \dfrac{2|x^T||y|}{|x^Ty|}$