Stabilisation de l'itération de Newton par correction automatique des erreurs d'arrondi

Philippe LANGLOIS

Journées FIABLE, 23 septembre 1999

IREMIA - Université de La Réunion
Projet CNRS/INRIA/ENSL ARENAIRE - E.N.S. Lyon

Plan

- Calcul de racines mal conditionnées de polynôme
- Limitations de l'itération de Newton
- Evaluation polynomiale
- CENA :

  Correction des Erreurs Numériques d'Arrondi

  ◁ Correction Linéaire

  ◁ Algorithme linéaire

  ◁ Automatisation
- Stabilisation et amélioration de la précision

Itération de Newton

Initialisation : choisir $X(0)$, $epsilon, alpha$

Itération : tant que TEST vrai

$$X(k+1) = X(k) - P[X(k)]/P'[X(k)]$$

Test(s) :

1. amélioration relative :

$$|X(k+1) - X(k)|/|X(k)| < epsilon$$

2. résidu absolu :

$$|P[X(k)]| < alpha$$

Convergence locale en X^* racine de multiplicité m :

$$\epsilon_n = |X(n) - X^*|$$

$$\epsilon_{n+1} = \frac{m-1}{m}\epsilon_n + O(\epsilon^2)$$

Quelques contraintes de la précision finie

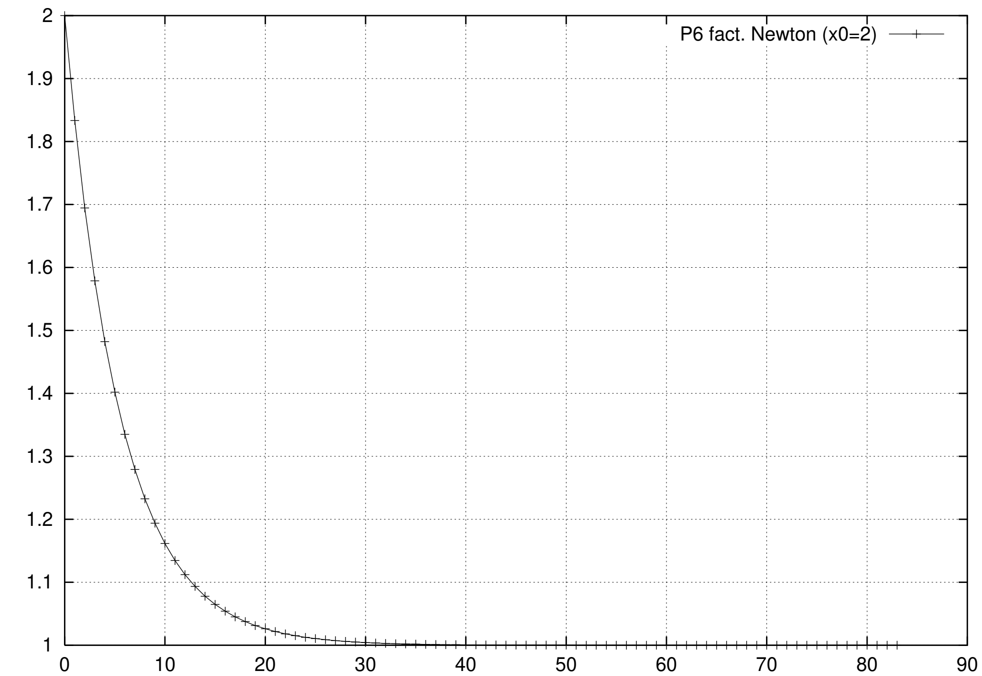
IEEE-754 simple précision, arrondi au plus près

- $\mathbf{u} = 2^{-24} \approx 5.96 \times 10^{-8}$,
- de 6 à 9 chiffres décimaux significatifs,
- $\lambda_n = 1.17 \times 10^{-38}$,
- $\lambda_d = 1.40 \times 10^{-45}$.

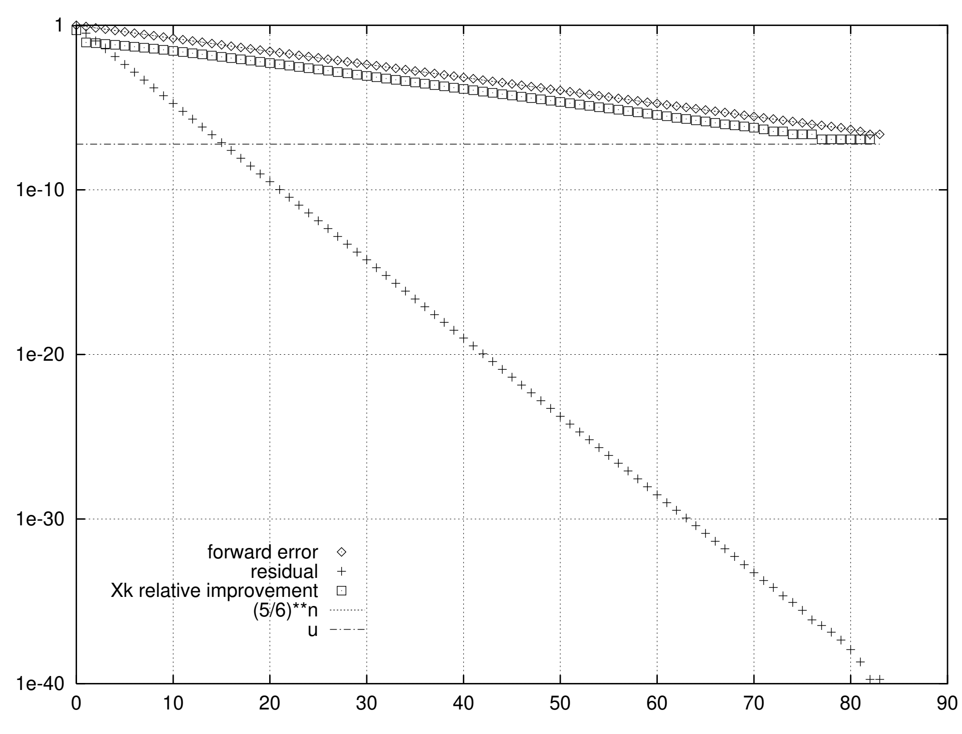
Evaluation et racines de $P_N(X) = (X - 1.0)^N$

- racine et coefficients exactement représentés,
- $epsilon = 1.0 \times 10^{-8} \gtrapprox \mathbf{u}$,
- $alpha = 1.0 \times 10^{-18} \ll \lambda_n$.

$P_6(X) = (X - 1.0)^6$ sous forme factorisée

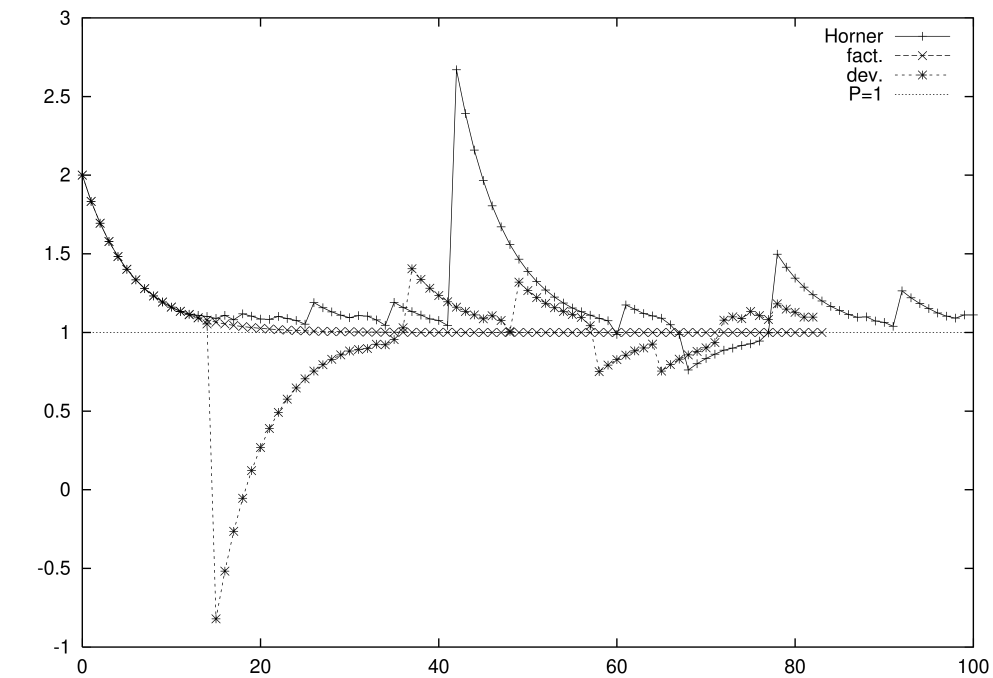


$$\widetilde{X^*} = X(82) = 1.000\_000\_24$$



Effet des erreurs d'arrondi

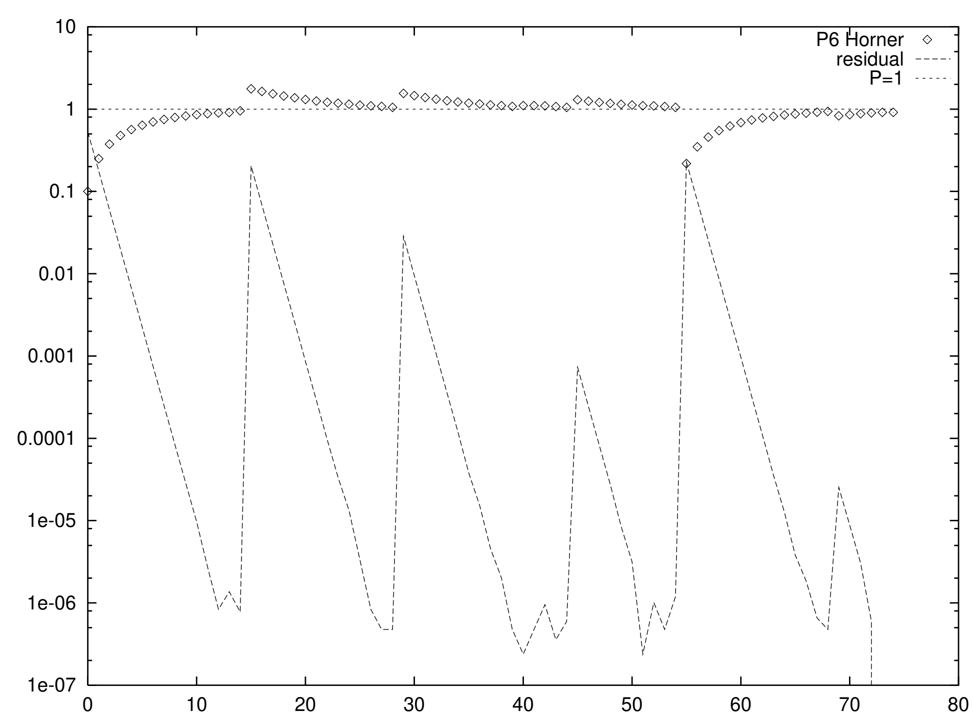
- P_6 sous formes développée et de Horner



  ▷ Horner : $P_6[X(100)] = 0.0$ pour $X(100) = 1.112\_084\_98$

  ▷ Développée : Breakdown pour $X(82) = 1.097\_851\_51$

- Influence de $X(0)$ (P_6 sous forme de Horner)



Précision, conditionnement et stabilité

[Wilkinson,63], . . . , [Chaitin-Chatelin & Frayssé, 96]

$P(X) = \sum_{i=0}^{N} a_i x^i$ avec $a_i \in \mathbb{R}$

$P(X^*) = P'(X^*) = \cdots = P^{(m-1)}(X^*) = 0$ et $P^{(m)} \neq 0$

- **Calculer les racines multiples d'un polynôme :**

  ▷ Précision limite atteignable [Stoer & Bulirsch,80] :

$$|X(k) - X^*| \leq \epsilon_{X^*} = \left(\frac{\underline{\delta} \cdot m!}{|P^{(m)}(X^*)|} \right)^{1/m},$$

$$\widetilde{P}[X(k)] = P[X(k)] + \delta(X(k)) \text{ et } |\delta(X(k))| \leq \underline{\delta},$$

  et ce indépendamment de l'algorithme de recherche de racine.

  ▷ Conditionnement relatif à la multiplicité m :

$$K_m(X^*, P) = \frac{1}{|X^*|} \left(\frac{m! \|P\| \|\underline{X^*}\|}{|P^{(m)}(X^*)|} \right)^{1/m}$$

  où $\underline{X^*} = (1, X^*, \cdots, (X^*)^N)^T$.

- **Evaluer un polynôme au voisinage de sa racine multiple est un problème bien conditionné :**

$$P(\widetilde{X^*}) = P(X^* + \epsilon) = O(\epsilon^m), \quad |\epsilon| \ll 1$$

  instabilité de l'itération de Newton et la précision de $\widetilde{X^*}$ *dépend de la qualité de l'algorithme d'évaluation*

Précision finie et évaluation polynomiale

- Forme de Horner est backward stable

  ... mais sa précision absolue est en $O(\mathbf{u})$:

  $\widetilde{P_6}(1+\epsilon) = \widetilde{Y}\,\widetilde{+}\,1.0$

  $\widetilde{P_6}(X) = 0.0$ ou $\widetilde{P_6}(X) \geq \mathbf{u}$ pour X voisin de X^*.

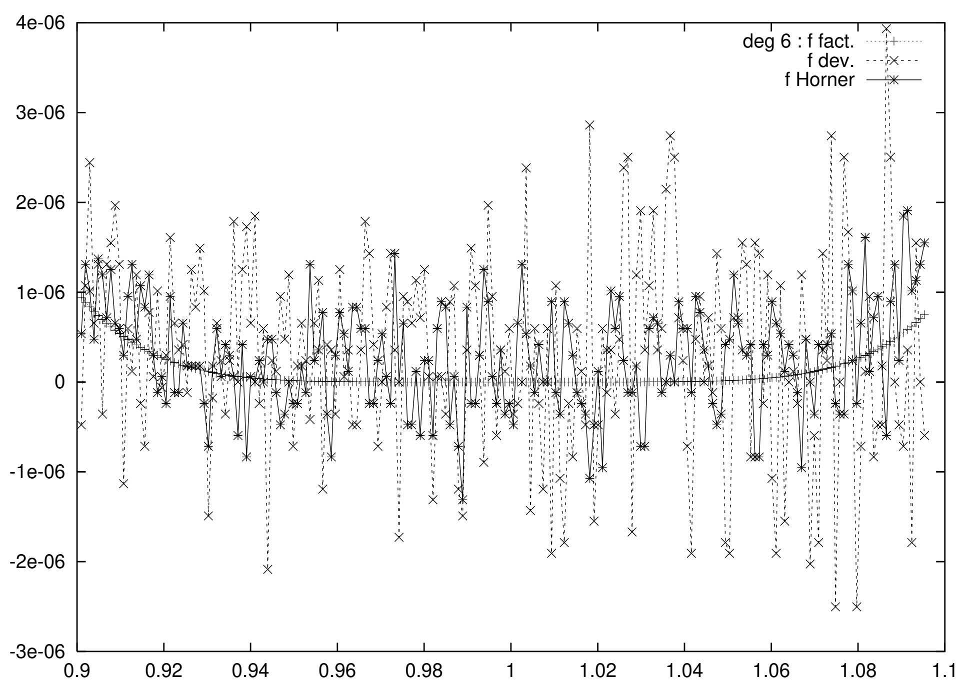
- La forme factorisée a une précision absolue en $O(\mathbf{u}^N)$

  au voisinage de la racine multiple :

  $\widetilde{P_6}(1+\epsilon) = \left((1\,\widetilde{+}\,\epsilon)\,\widetilde{-}\,1.0\right)^6$,

  $\widetilde{P_6}(1+\epsilon) = 0.0$ pour $|\epsilon| < \mathbf{u}$ et

  $\widetilde{P_6}(1+\epsilon) = O(\mathbf{u}^6)$ sinon.



Itération de Newton et *Running Error Analysis*

Arréter l'itération quand $P[X(k)]$ n'est plus significatif

[Kahan,63], [Peters et Wilkinson,71], [Igarashi,84]

- l'itération est arrétée avant qu'elle ne devienne instable,

▷ n'améliore pas la précision de l'itéré.

Initialisation : choisir $X(0)$

Test(s) :

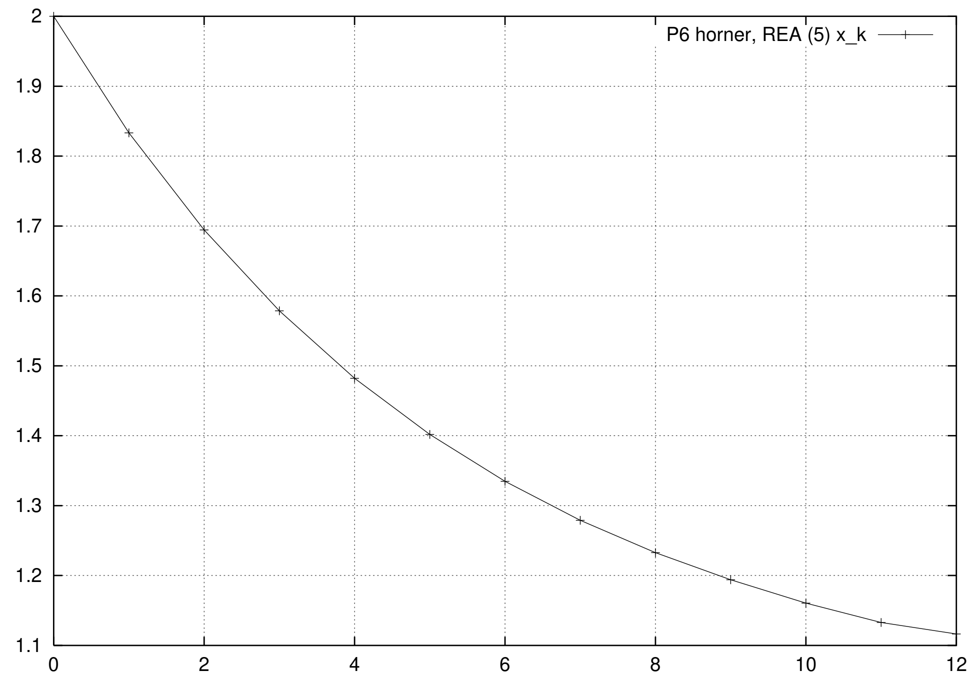
3. Si REB majore la précision effective de $P[X(k)]$, arrêt si

$$C \times |P[X(k)]| < REB(k)$$

Horner et REB associée :

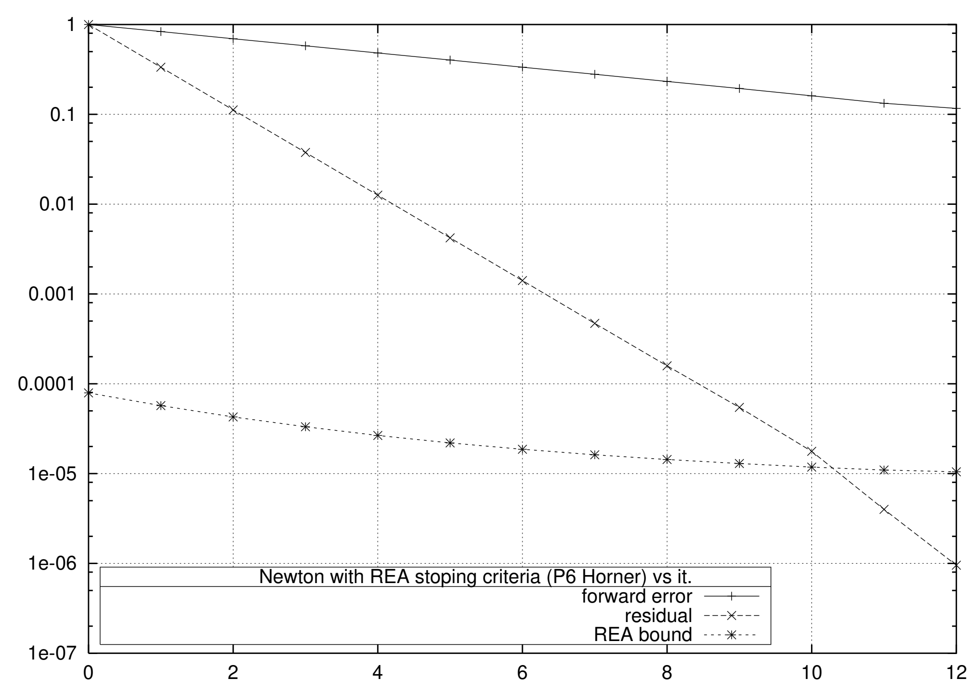
```
P[N] = a(N)
REB = 0
for I = N-1 , 0, pas=-1
  P[i]=x*P[i+1]+a(i)
  REB = |x|*REB + |x|*|P[i+1]| + |a(i)|
end
P=P(0)
REB = REB * u
```

$P_6(X)$ et test d'arrêt REA



$$\widetilde{X^*} = X(11) = 1.132\_770\_06$$

$$|\widetilde{X^*} - 1.0| = O((\mathbf{u})^{1/8})$$



Correction Linéaire

Soient $f : \mathbb{F}^n \to \mathbb{R}^m$ et $\widetilde{f} : \mathbb{F}^n \to \mathbb{F}^m$ $(m = 1)$.

On calcule $\widetilde{x_N} = \widetilde{f}(X)$ pour $X = (x_1, \ldots, x_n) \in \mathbb{F}$.

- Erreur élémentaire δ_k associée à x_k $(n + 1 \leq k \leq N)$:
  $\widetilde{x}_i, \widetilde{x}_j, \widetilde{x_k} \in \mathbb{F}$, $(1 \leq i \leq j < k \leq N)$
  $\circ \in \{+, -, \times, / , \sqrt{\ }\}$ et $\widetilde{\circ}$ correspondant

$$\widetilde{x_k} = \widetilde{x}_i \widetilde{\circ} \widetilde{x}_j = \widetilde{x}_i \circ \widetilde{x}_j + \delta_k = x_k + \delta_k$$

- Forward erreur globale :

$$FE(X) = \widetilde{f}(X, \delta) - \widetilde{f}(X, 0)$$

- Linéarisation par rapport aux E.A. élémentaires :

$$FE(X) = \sum_{k=n+1}^{N} \frac{\partial \widetilde{f}}{\partial \delta_k}(X, \delta) \cdot \delta_k - E_L = \Delta_L - E_L$$

- Correction Linéaire :

$$\overline{f(X)} = \widetilde{f}(X, \delta) + \widetilde{\Delta_L} \text{ avec } \widetilde{\Delta_L} = \Delta_L - E_C$$

Algorithme Linéaire

Définition 1 *Un* algorithme linéaire sur $\mathbb{F}$ *est un algorithme qui utilise les opérations* $\{+, -, \times, /, \sqrt{}\}$ *et tel que*
- au moins un des deux opérandes de chaque multiplication,
- le diviseur de chaque division,
- l'opérande de chaque racine carrée
appartient à $\mathbb{F}$*.*

Théorème 1 *Soit* $\widetilde{x_N}$ *le résultat calculé de* $f(X)$ *par un algorithme linéaire sur* $\mathbb{F}$*. La forward erreur globale* $\widetilde{x_N} - f(X)$ *est une fonction linéaire des erreurs d'arrondi élémentaires.*

Corollaire 1 *Lorsqu'un algorithme linéaire est appliqué pour calculer* $f(X)$*, le résultat corrigé* $\overline{f(X)}$ *n'est entaché que de l'erreur* E_C *de calcul du terme correcteur,* i.e., *.*

$$\overline{f(X)} = f(X) + E_C. \tag{1}$$

Exemple 1 *Algorithmes de sommation (ordre naturel), produit scalaire, résolution de systèmes linéaires triangulaires,* ***évaluation polynomiale par la méthode de Horner****, ...*

CENA : automatisation de la correction

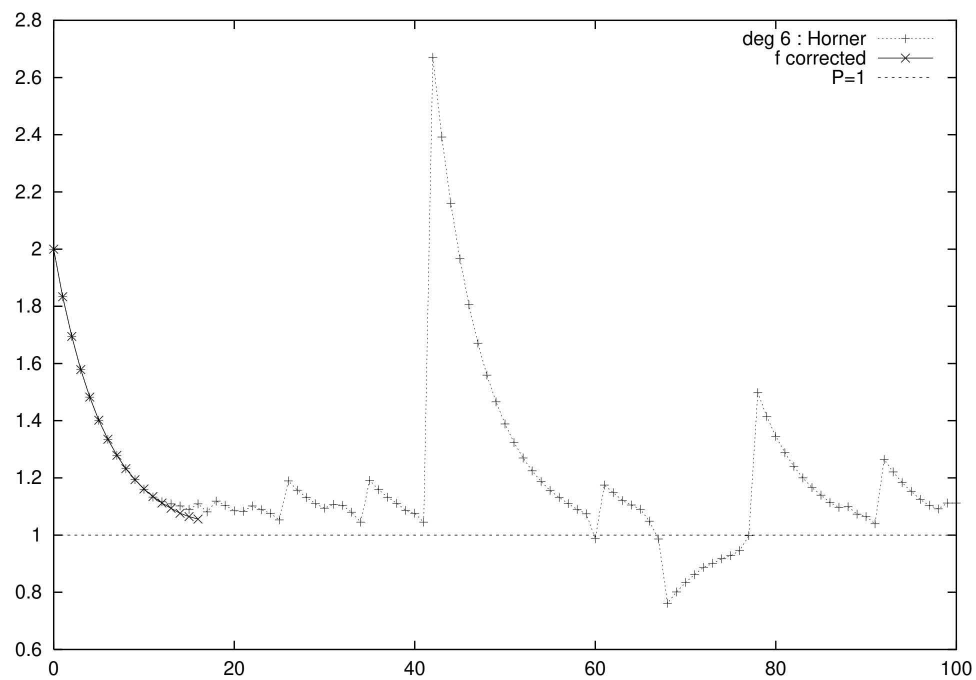
$$\overline{f(X)} = \widetilde{f(X)} \,\widetilde{+}\, \left(\frac{\partial \widetilde{f}}{\partial\, \delta_k}(X,\delta)\right)^T \widetilde{\cdot} \left(\widetilde{\delta_k}\right)$$

- Dérivées partielles
  - différentiation automatique en mode inverse
  - majoration de l'erreur de calcul par Running Error Bound
- Erreurs élémentaires $\delta[\circ]$
  - $\circ \in \{+, -, \times, \}$: $\delta[\circ]$ appartiennent à $\mathbb{F}$ et sont calculables sur $\mathbb{F}$;
  - $\circ \in \{/, \sqrt{\ }\}$: une estimation de $\delta[\circ]$ est calculable sur $\mathbb{F}$ ainsi qu'une majoration de l'erreur d'approximation.
- Produit scalaire
  - majoration *a priori* existe [Nativel;98]
  - majoration *a posteriori* par Running Error Bound

Erreurs d'arrondi élémentaires

$\circ$	$\delta[\circ, (x_i, x_j)],\ e(x_i) \leq e(x_j) \qquad t_{\pm} = t_{\times} = T, t_{/} = 10T, t_{\surd} = 20T$	β_k	$t_{\delta[\circ]}/t_{\circ}$
$\pm$	$\tilde{\mp}\left[x_i \tilde{-} (x_i \tilde{+} x_j)\right]$	0	2
$\times$	$(x_i \tilde{\times} x_j) \tilde{-} (x_i^U \tilde{\times} x_j^U)\ \left[(x_a^U \tilde{\times} x_j^L) \tilde{+} (x_i^L \tilde{\times} x_j^U)\right]\ \tilde{-} (x_i^L \tilde{\times} x_j^L)$, with $x_k^U = \left[x_k \tilde{-} (x_k \tilde{-} C)\right] \tilde{+} (x_k \tilde{\times} C)$, and $x_k^L = x_k \tilde{-} x_k^U$, $(k = i, j)$. $C = 2^{12} + 1$ for IEEE-754 single precision	0	≈ 13
$/$	$\left\{\left[(x_i \tilde{/} x_j) \tilde{\times} x_j\right] \tilde{-} x_i \tilde{-} \delta[\times, (x_i \tilde{/} x_j, x_j)]\right\} \tilde{/} x_j$	$x_i \tilde{/} x_j$	3
$\sqrt{\ }$	$(\widetilde{\sqrt{x_i}} \tilde{\times} \widetilde{\sqrt{x_i}}) \tilde{-} x_i \tilde{+} \delta[\times, (\sqrt{x_i}, \sqrt{x_i})]$	$\frac{5}{2}\widetilde{\sqrt{x_i}}$	≈ 1.5

$P_6(X)$ avec correction P

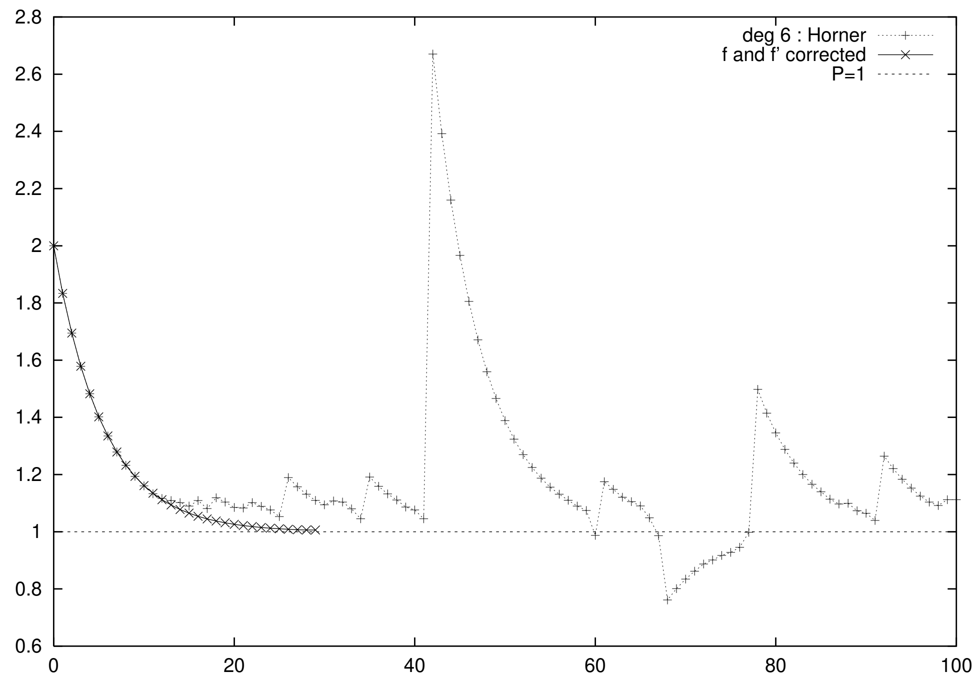


▷ Breakdown : zéro de $\widetilde{P'}$ pour

$$\overline{X^*} = \overline{X(16)} = 1.050\_682\_31$$

$$\widetilde{X(16)} = 1.053\_181\_77$$

$P_6(X)$ avec correction de P et P'



▷ L'itération est stabilisée

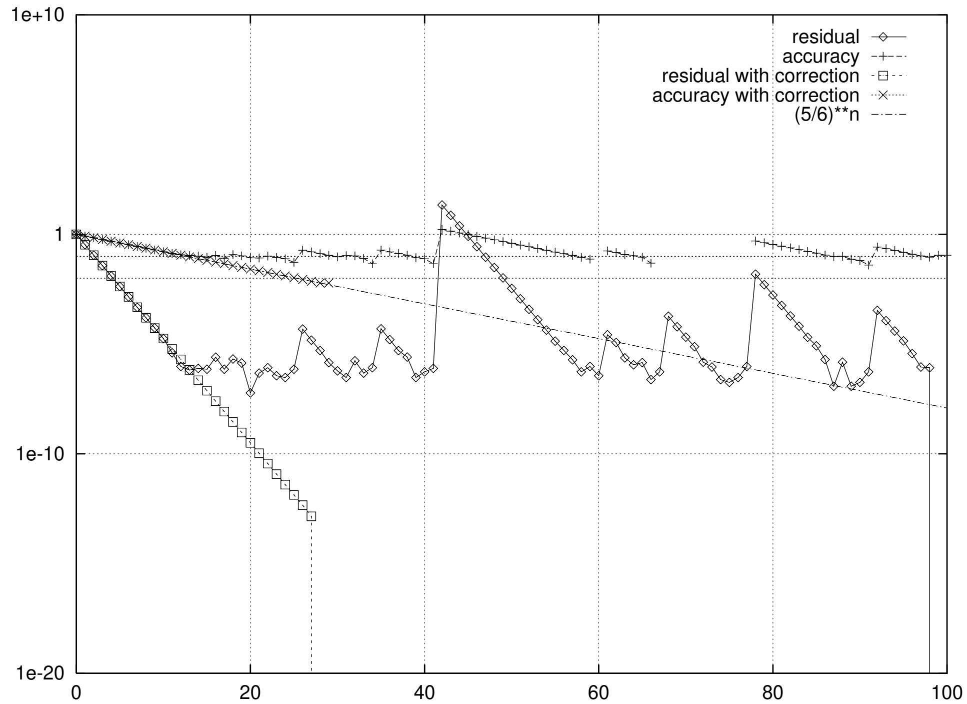
$$\overline{X^*} = \overline{X(29)} = 1.006\_008\_39$$

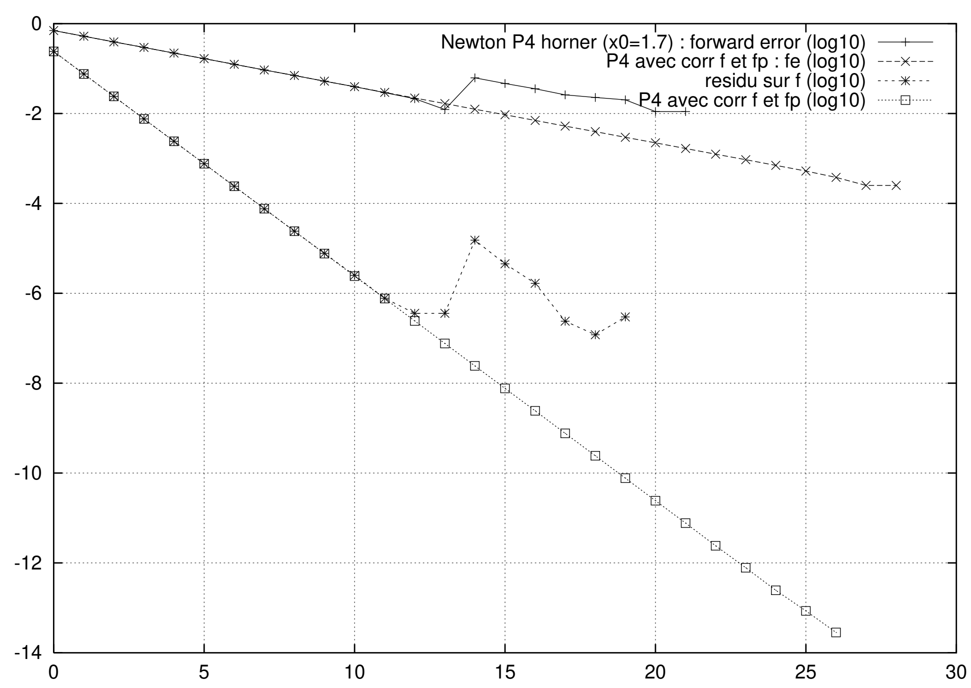
▷ Précision meilleure que la limite pour évaluation en SP

$$|\overline{X^*} - 1.0| = O((\mathbf{u})^{1/3})$$

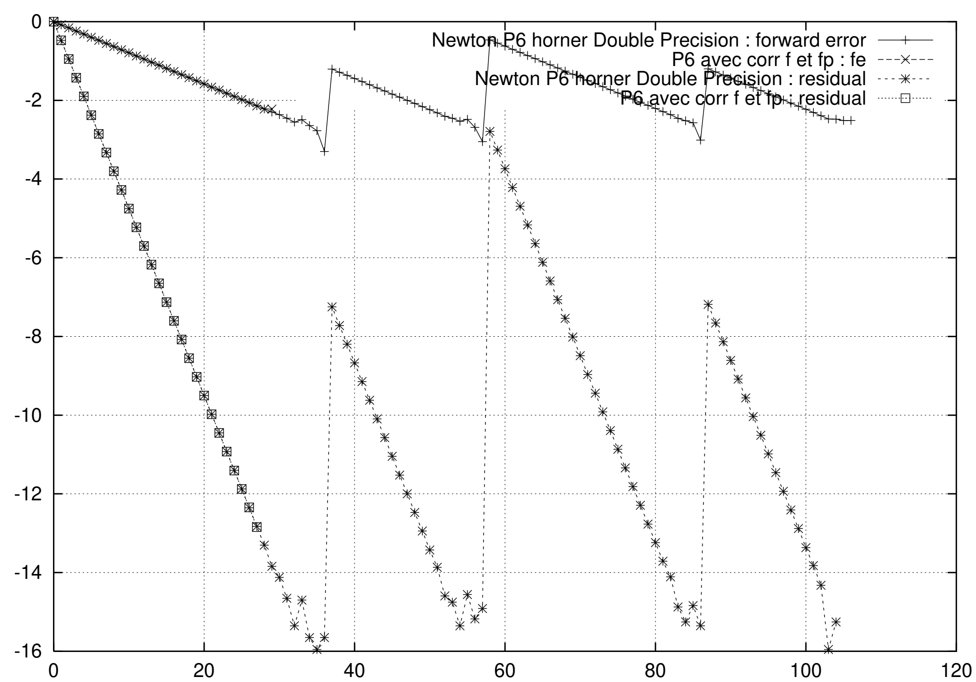
$P_6(X)$ et $P_4(X)$ avec correction de P et P'

▷ Ordre de multiplicité retrouvé :





Comparaison avec la double précision



Conclusion et perspectives

- stabilisation
- amélioration de la précision
- exemple de correction de variable sensible
  - ▷ correction du résultat final [CRAS,98],[CRAS,99]
- singularité *vs* limites théoriques méthodes différentielles

- automatisation pas gratuite
- extension aux fonctions élémentaires
- auto-validation sur algorithmes non linéaires
- ...

Références

[Chaitin-Chatelin & Frayssé, 96] **Françoise Chaitin-Chatelin & Valérie Frayssé**
Lectures on Finite Precision Computations,
SIAM,1996.

[CRAS,98] **Philippe Langlois & Fabrice Nativel**
Reduction and Bounding of the Rounding Error in Floating Point Arithmetic,
C.R. Acad. Sci. Paris, Série 1, 327, pp.781–786, 1998.

[CRAS,98] **Philippe Langlois & Fabrice Nativel**
When Automatic Linear Correction of Rounding Errors is Exact,
C.R. Acad. Sci. Paris, Série 1, 328, pp.543–548, 1999.

[Igarashi, 84] **Masao Igarashi**
A termination criterion for iterative methods used to find the zeros of polynomials,
Mathematics of Computation, 42, pp.165–171, 1984.

[Nativel, 98] **Fabrice Nativel**
Fiabilité numérique et précision finie,
Thèse de Doctorat de l'Université de La Réunion, Décembre 1998.

[Peters & Wilkinson,71] **G. Peters & J. H. Wilkinson**
Practical Problems Arising in the Solution of Polynomial Equations,
J. Inst. Maths Applics, 8, pp.16–35, 1971.

[Pichat & Vignes, 93] **Michèle Pichat & Jean Vignes**
Ingénierie du contrôle de la précision des calculs sur ordinateur,
Technip, Paris, 1993.

[Wilkinson,63] **J. H. Wilkinson**
Rounding Errors in Algebraic Processes,
Prentice-Hall, Englewood Cliffs, NJ, 1963.