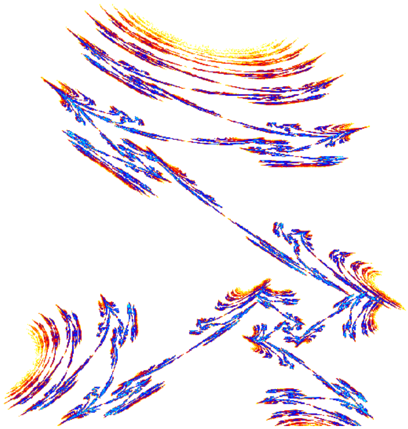
THEORIE

DES ALGORITHMES ÉVOLUTIONNAIRES



Evelyne LUTTON

Equipe APIS - INRIA Saclay - Ile-de-France - Evelyne.Lutton@inria.fr

http ://complex.inria.fr/

THÉORIE DES AEs

Questions

- Convergence.
- Fonctions “déceptives.”
- Choix et ajustement des paramètres.
- Représentation/Codage.

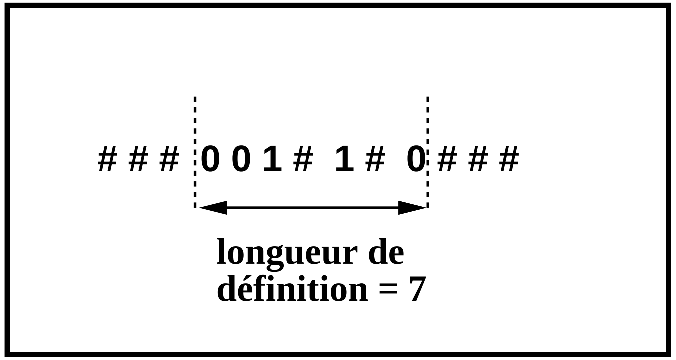
Deux grandes approches

- Théorie des Schémas *(Holland, 1975)*
- Modélisation Markovienne *(depuis 1987)*

THÉORIE DES SCHÉMAS

Intuition : les similarités entre codes guident la recherche.

Notion de schéma : le schéma $\#001$ représente le sous ensemble : $\{0001, 1001\}$



– **Nombre d'allèles fixés :** $\mathcal{O}(H)$

– **Longueur de définition :** $\delta(H)$

LE THEORÈME DES SCHÉMAS, pour des populations de taille infinie. : ***(Holland, 1975)***

$$E(m(H,t+1)) \geq m(H,t)\frac{\overline{f(H)}}{\bar{f}}[1 - p_c\frac{\delta(H)}{l-1} - \mathcal{O}(H)p_m]$$

NOTION DE DÉCEPTIVITÉ

Building blocks = "bons" schémas ayant des $\mathcal{O}(H)$ et des $\delta(H)$ faibles.

- ***AG-facile* :** l'optimum global de f appartient à l'intersection des building blocks.
- ***AG-difficile* :** l'intersection des building blocks est toujours un optimum local.
- ***Déceptivité statique*** *(Goldberg 89)* : attraction de l'AG vers les optima d'une fonction f' :

$f'(i) = E(f(i'))$ où i' peut être atteint à partir de i par mutation et croisement.

Si l'optimum global de f' et de f diffèrent la fonction est déceptive.

MODÉLISATION MARKOVIENNE

Le passage d'une génération à la suivante peut être considéré comme un processus stochastique dans un espace d'états fini. $\longrightarrow$ **Chaîne de Markov**

- ***Goldberg et Segrest, 1987 :***
  – Chromosomes de longueur 1
  – Population de taille finie
  ==> dérive génétique

- ***Horn, 1993 :***
  – Niches écologiques

- ***Davis et Principe, 1991 :***
  – chromosomes de longueur > 1
  ==> décroissance de la probabilité de mutation

- ***Nix et Vose, 1992 :***
  – la population croît pendant l'évolution

- ***Cerf, 1993.***

Les résultats sur la convergence

Convergence “dans le cas le pire” :

Résultats globaux (simples).

– Théorie des Schémas,

– Modélisations Markoviennes,

– “The genetic algorithm fractal” (Juliany & Vose).

Fitness contrôlé :

Les résultats sont plus précis grâce à des hypothèses restrictives sur la fonction de fitness :

– NK Landscapes,

– Stratégies d’évolution sur des modèles de sphères,

– Analyse de régularité fractale.

Paysages irréguliers et fractals

L'irrégularité est une source de difficultés pour les AEs

- Existe-t'il un lien entre irrégularité et performance des AEs ?
- Comment rendre les AEs plus efficaces sur des paysages de fitness irréguliers ?

Caractérisation uniforme de la régularité de la fonction de fitness

Soient (X, d_X) **et** (Y, d_Y) **deux espaces métriques.**

$F : X \rightarrow Y$ **est une fonction Höldérienne d'exposant** $h \geq 0$

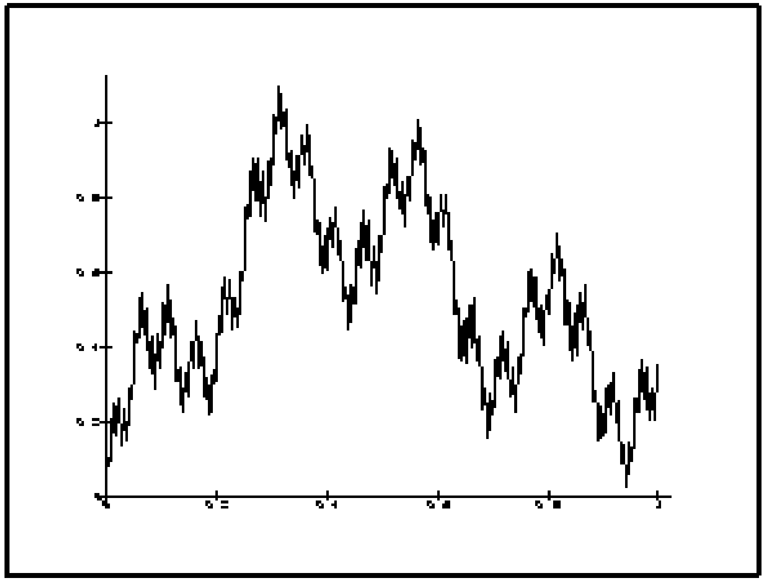
si $\forall x, y \in X$ **tel que** $d_X(x, y) < 1$ $\quad d_Y(F(x), F(y)) \leq k.d_X(x, y)^h$

∞ **pour une constante** $k > 0$.

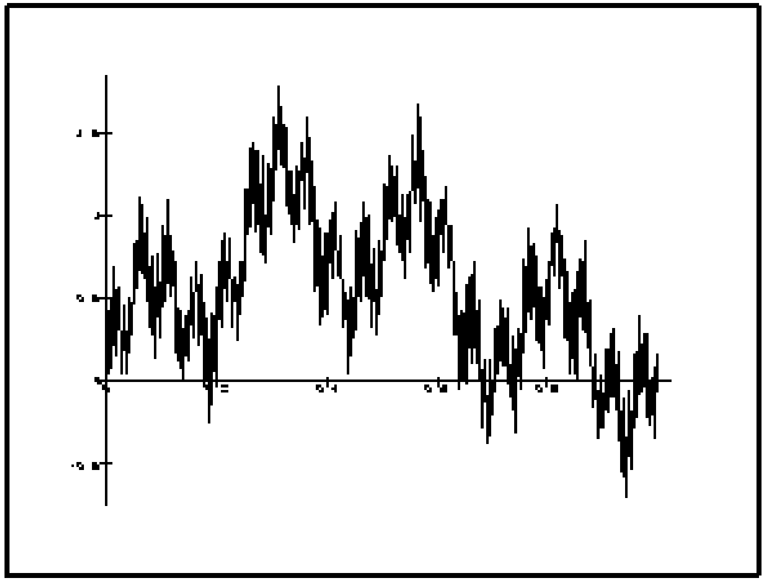
- **Si** F **est Höldérienne d'exposant** h**, elle est Höldérienne d'exposant** h' **pour tout** $h' \in (0, h]$.
- **Une fonction A Höldérienne est toujours continue, mais non nécessairement différentiable.**

Fonctions de Weierstrass

$$W_{b,s}(x) = \sum_{i=1}^{\infty} b^{ih} \sin(b^i x) \quad \textbf{avec } b > 2 \ \ \textbf{et } 0 < h < 1$$



Dimension $s = 1.5$ **(Hölder** $h = 0.5$**)**



Dimension $s = 1.7$ **(Hölder** $h = 0.3$**)**

Fonction + topologie génétique = fitness landscape

La mesure de régularité dépend de la métrique définie sur l'espace de recherche.

Pour les AEs : l'irrégularité apparente dépend de la topologie "génétique."

$\rightarrow$ Le design des opérateurs génétiques est extrêmement important.

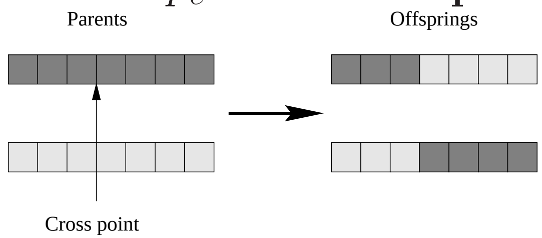
L'algorithme génétique canonique

Fonction de fitness : $f : \Omega^l = \{0,1\}^l \rightarrow \mathbb{R}^+$

– **Sélection proportionnelle :**

$$P(i) = \frac{f(i)}{\sum_{j=1}^{N} f(j)}$$

– **Crossover à un point** avec probabilité p_c sur un couple d'individus :



– **Mutation** avec une probabilité fixée faible, p_m.

Mesure de difficulté pour un paysage de fitness

– ***Static deception (Goldberg 89) :*** **L'AG est attiré vers les optima de f' :**

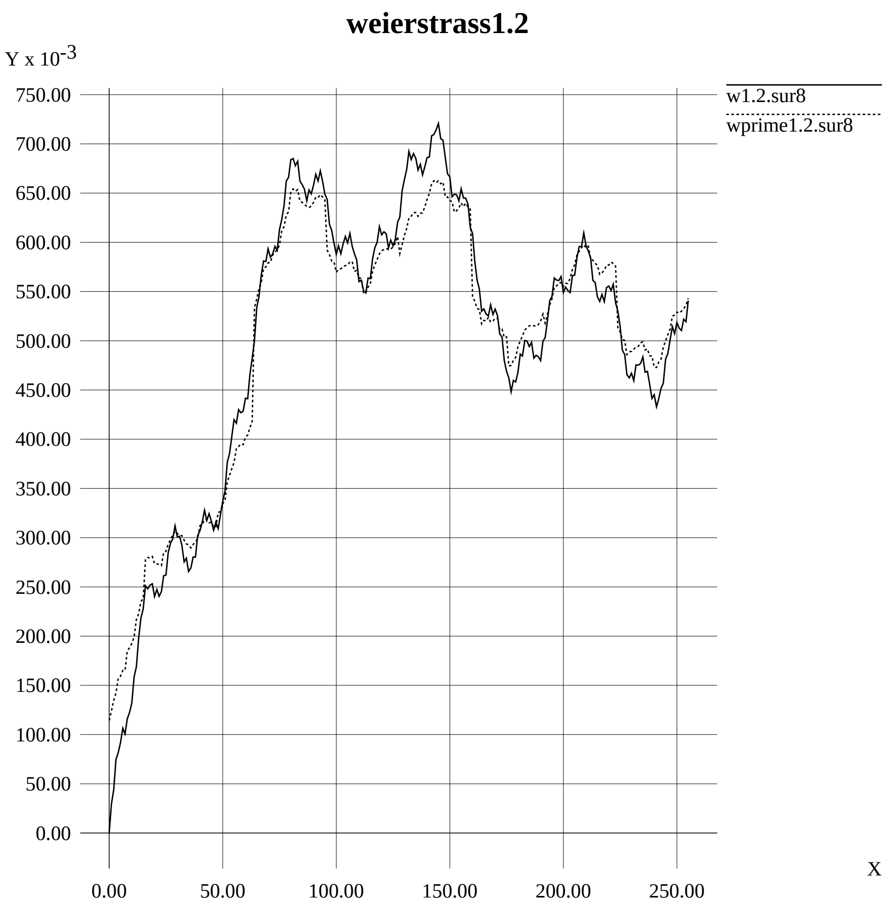
$$f'(i) = E(f(i'))$$

où i' est atteint a partir de i par une mutation ou un crossover.

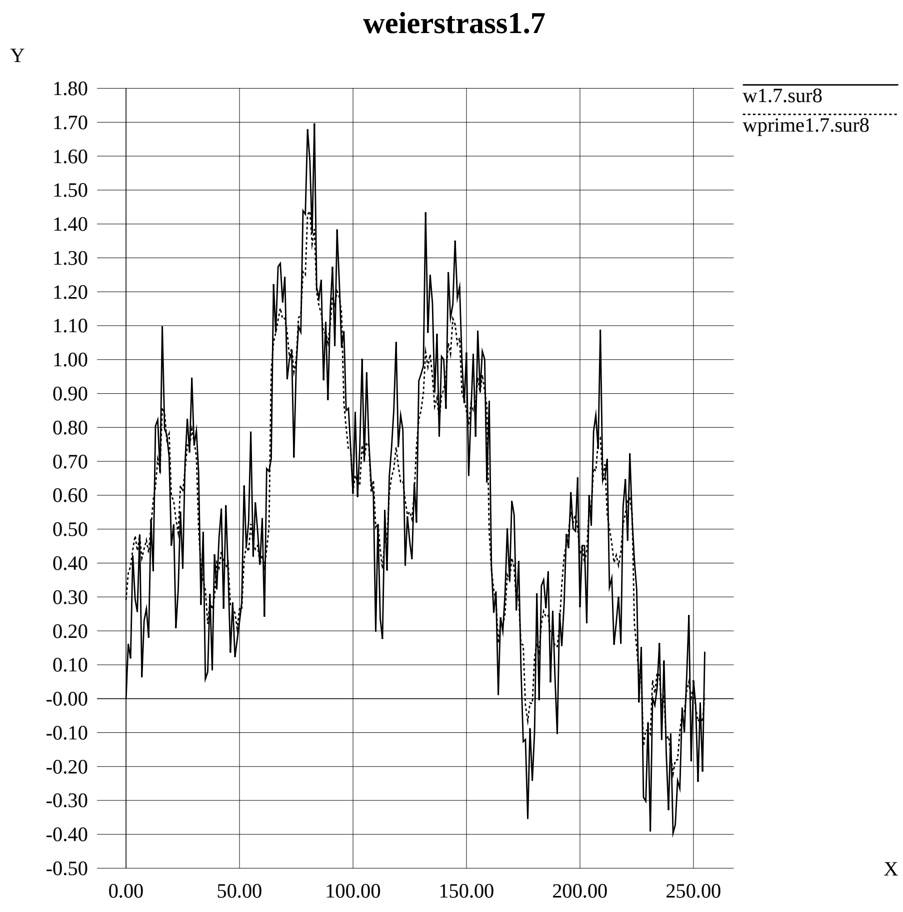
Si l'optimum global de f' et de f diffèrent, la fonction est "trompeuse" ($\simeq$ AG-difficile)

f' peut être calculée à partir de f, p_m et p_c grâce à une décomposition sur une base de Walsh.

Fitness ajusté pour les fonction de Weierstrass



Hölder $h = 0.8$



Hölder $h = 0.3$

Fitness ajusté et taux de progrès

$$\Delta f = |f(x) - f'(x)| = |f(x) - E_{Voisins``Genetiques''}(f(x))|$$

Δf est un gain espéré (– ou perte désespérée !) pour la fonction de fitness en une application des opérateurs génétiques.

Pour un (1+1)ES sur un espace de recherche continu, Δf est directement lié au taux de progrès.

Fonctions Höldériennes

Si f est l'échantillonnage d'une fonction Höldérienne F sur $[0,1]$, d'exposant h et de constante k :

$$\forall x \in \{0,1\}^l, \qquad f(x) = F(\frac{I(x)}{2^l})$$

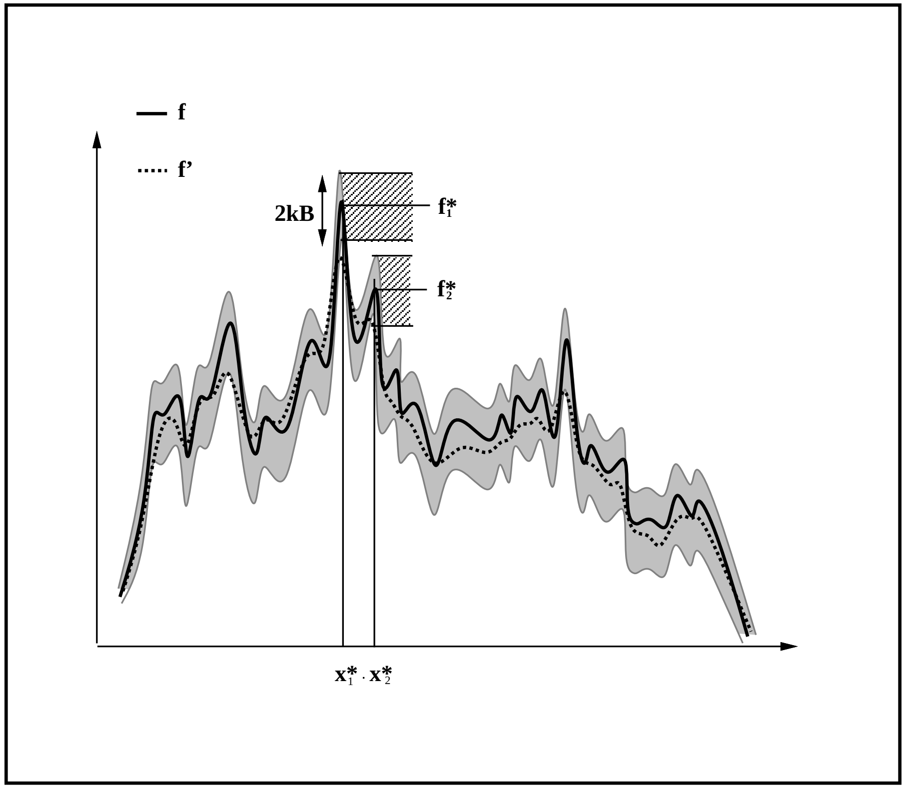
$I(x) \in [0, 2^l - 1]$ **est l'entier dont la décomposition binaire est** x

$$\forall x \in \{0,\ldots,2^l-1\} \quad |f(x) - f'(x)| \leq k.B(p_m, p_c, l, h)$$

avec

$$\begin{aligned} B(p_m,p_c,l,h) = \quad & \frac{p_c}{l-1} 2^{-h} \left[\frac{2^{-l(h+1)}-1}{2^{-(h+1)}} + \frac{(1-2^{l-h})(2^{-hl}-1) - l2^{-hl}(1-2^{-h})}{(2^{-h}-1)^2} \right] \\ & + p_m \frac{2^{-h}}{(2^{-h}-1)^2} \left[1 + 2^{-hl}(l2^{-h} - l - 1) \right] \end{aligned}$$

Majoration de $\Delta f = |f(x) - f'(x)|$



- B **décroît selon** h.
- B **croît selon** p_m **et** p_c.
- B **croît selon** l **pour** l **petit, atteint un maximum en** l_{max}**, puis décroît pour** $l > l_{max}$.

Coefficients de régularité bit-à-bit : caractérisation fondée sur la distance de Hamming.

Définition : **soit f une fonction définie sur** Ω^l **:**

$$\forall q \in \{0, \ldots, l-1\},\ C_q = \sup_{x \in \Omega^l} \{|f(x) - f(x'_{l-q-1})|\}$$

avec x'_{l-q-1} **et** x **différents uniquement selon le bit de position** $(l-q-1)$

Théorème :

Soit f **une fonction définie sur** Ω^l **dont les coefficients de régularité bit-à-bit sont** $(C_q)_{q \in \{0,\ldots,l-1\}}$**.**
Alors $\forall x \in \Omega^l$ **:**

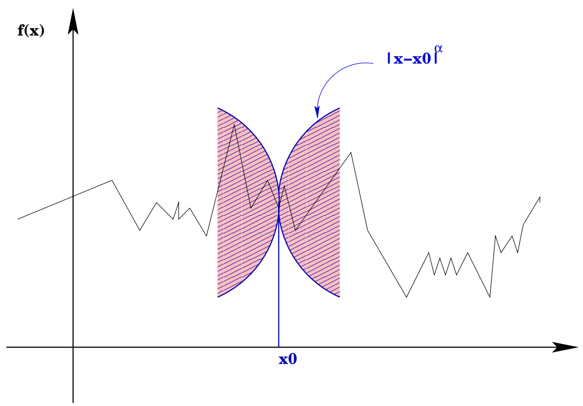
$$|f(x) - f'(x)| \leq \frac{p_c}{l-1} \sum_{q=0}^{l-1} C_q \left(\frac{1 + 2^q(q-1)}{2^q} \right) + p_m \sum_{q=0}^{l-1} C_q(q+1)$$

Régularité locale

Soit $\alpha \in (0,1)$**,** $\Omega \subset \mathbf{R}$**.**

$f \in C_l^\alpha(\Omega)$ **ssi** $\exists\ k : \forall x, y \in \Omega : |f(x) - f(y)| \leq k|x-y|^\alpha$

$$\alpha_l\left(f, x_0, \rho\right) = \sup\left\{\alpha : f \in C_l^\alpha\left(B\left(x_0, \rho\right)\right)\right\}$$



L'exposant de Hölder local **d'une fonction continue** f **en** x_0 **est :**

$$\alpha_l\left(f, x_0\right) = \lim_{\rho \to 0} \alpha_l\left(f, x_0, \rho\right)$$

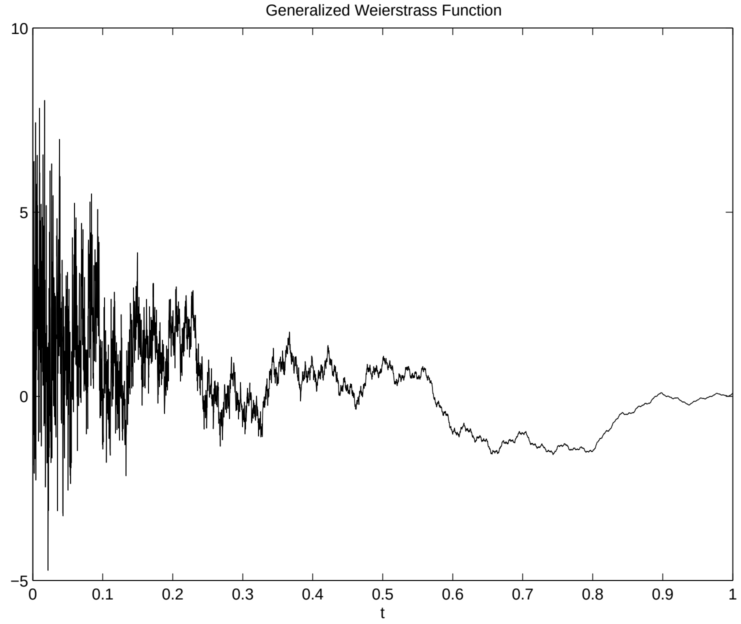
Expérience : influence de la régularité locale

L performance d'un ES est-elle affectée par un changement de régularité locale ?

Fonction de Weierstrass généralisée :

$$GW_{b,h}(x) = \sum_{i=1}^{\infty} b^{-ih(x)} sin(b^i x)$$

with $b \geq 2$ **and** $0 < h(x) < 1$



Si h **est différentiable, l'exposant de Hölder local de** $GW_{b,h}$ **est** $h(x)$ **en tout** x**.**

Fonctions-test

Composante lisse + irrégularité normalisée et contrôlée sur $[-0.5, 0.5]$:

$$f(x) = 2 - 4x^2 - |NW_{b,h}(x)|$$

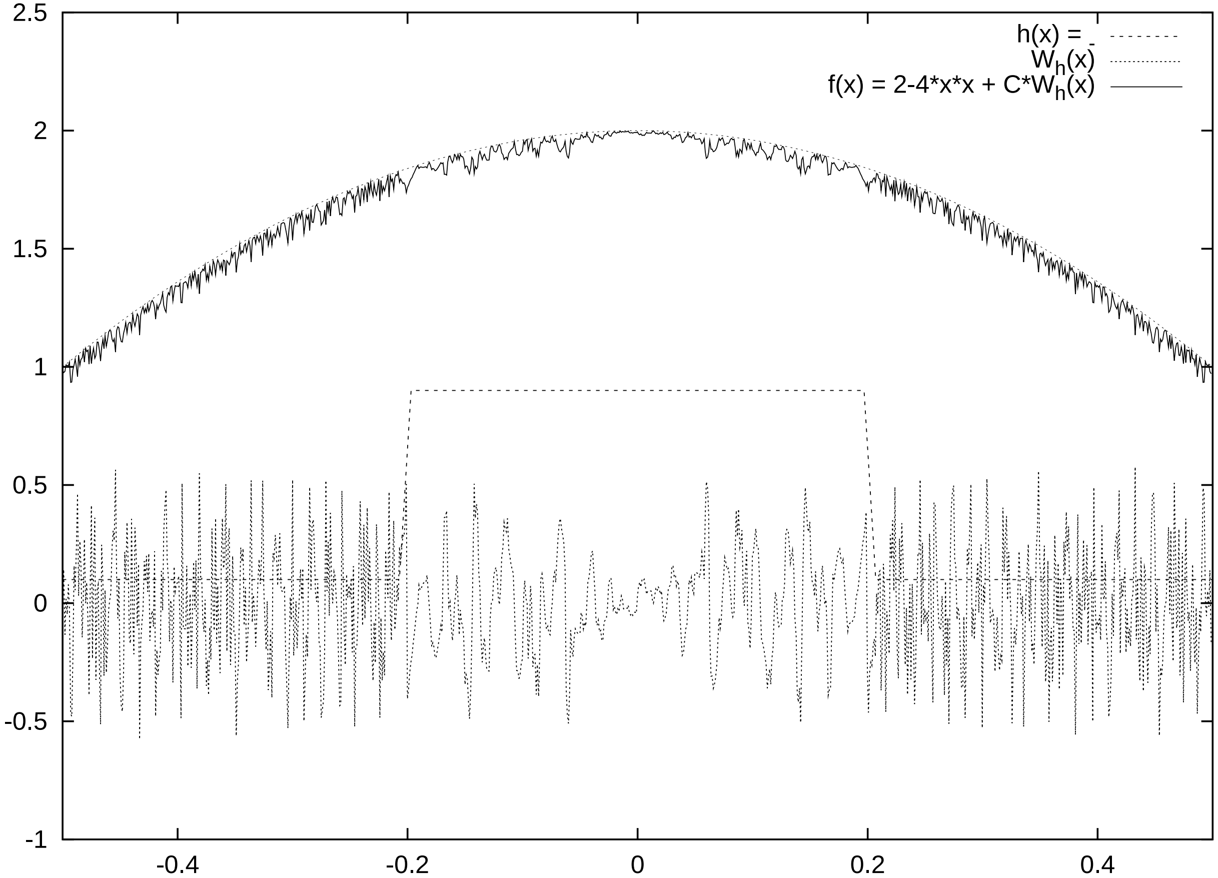
1. **$N(x)$, cas favorable : les zones irrégulières on un fitness peu élevé**

$$\begin{array}{lll} h(x) = 0.9 & \textbf{si} & x \in [-0.2, 0.2] \\ h(x) = 0.1 & \textbf{sinon} & \end{array}$$

2. **$U(x)$, cas défavorable : les points les plus irréguliers sont au voisinage de l'optimum global**

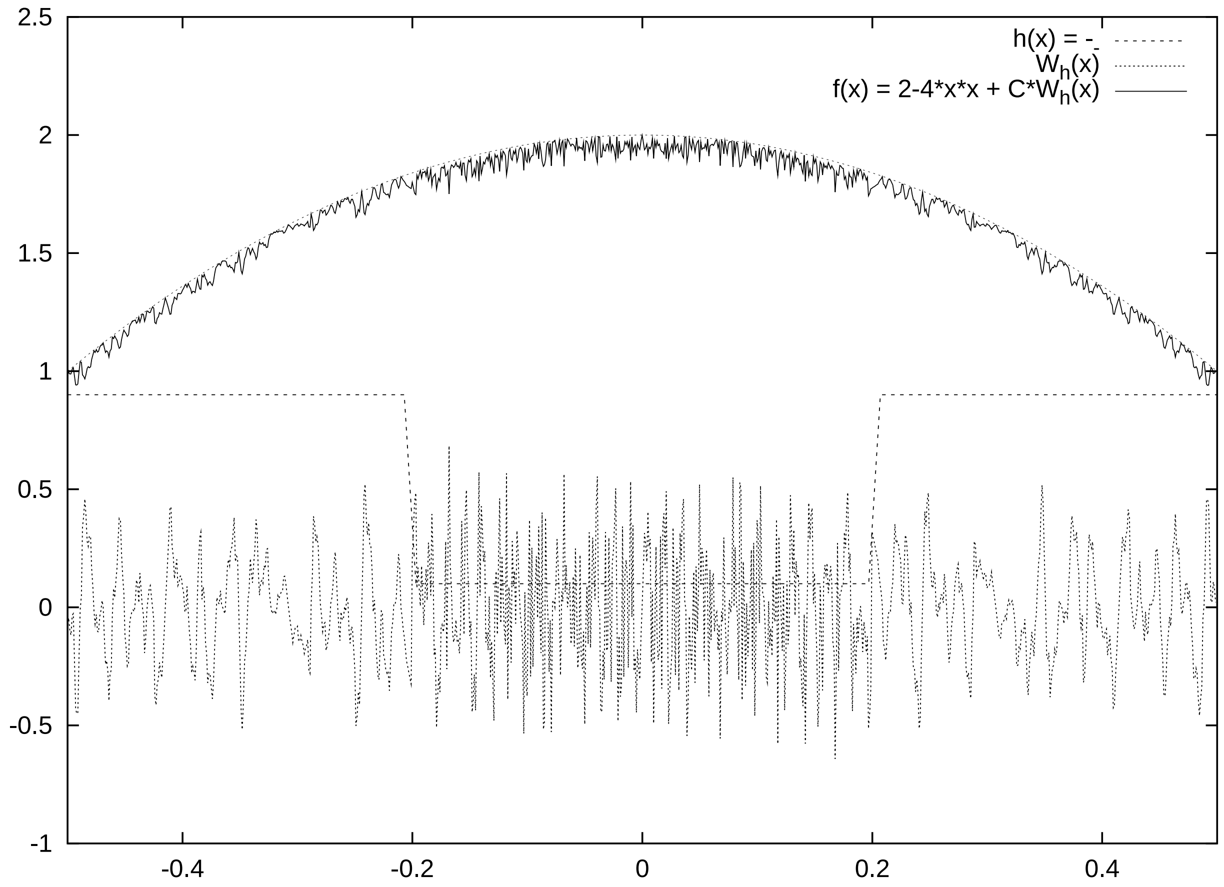
$$\begin{array}{lll} h(x) = 0.1 & \textbf{si} & x \in [-0.2, 0.2] \\ h(x) = 0.9 & \textbf{sinon} & \end{array}$$

Fonction-test N



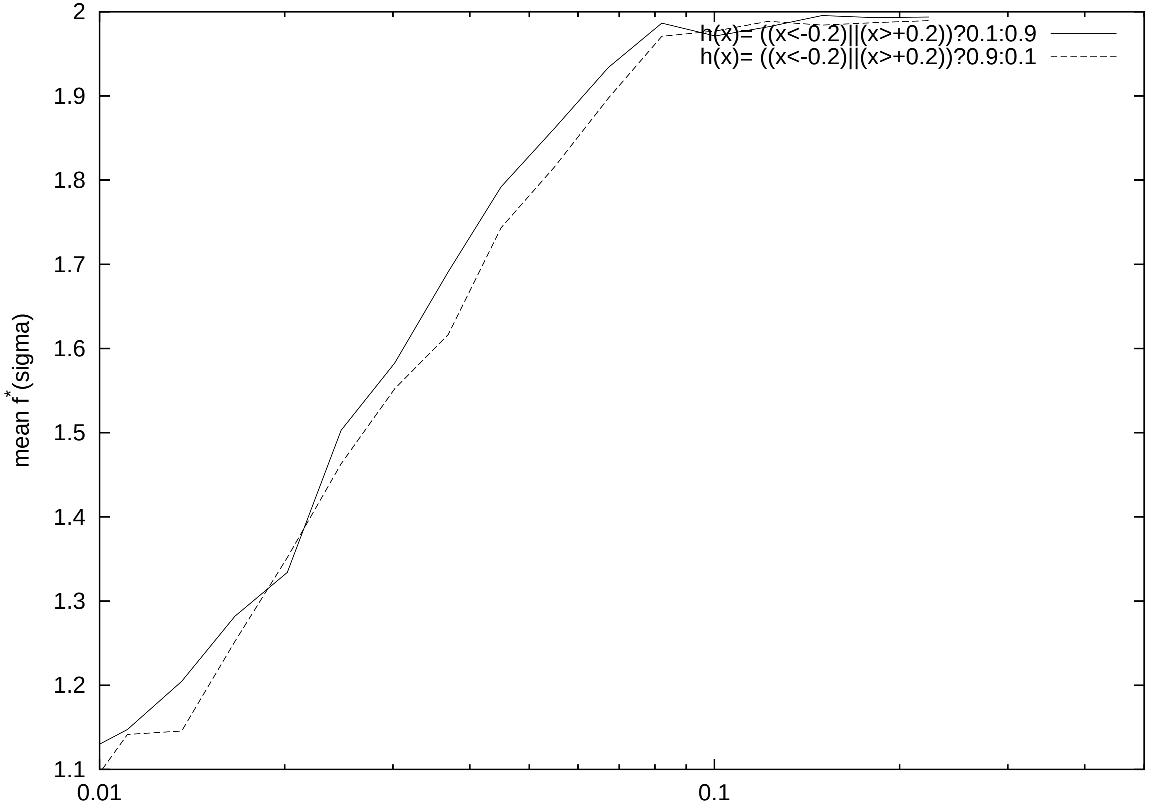
N(x) : le profil de régularité "n"

Fonction-test U



U(x) : le profil de régularité "u"

Test : (1+1)ES avec mutation uniforme de rayon σ



Moyenne des meilleurs fitness au bout de 300 generations d'un (1+1)ES pour U et N en fonction de σ.

Analyse théorique d'un ES à mutation uniforme

Pour une mutation uniforme de rayon σ, f' est le fitness espéré sur un disque de rayon σ.

$$f'(x) = \frac{1}{2\sigma} \int_{x-\sigma}^{x+\sigma} f(t)dt$$

Comme f est localement Hölder, nous avons pour tout x : $\forall t \in B(x,\sigma) \quad |f(x) - f(t)| \leq C_x |x-t|^{\alpha(x)}$

$|f(x) - f'(x)|$ est donc majoré en fonction de l'exposant de Hölder local $\alpha(x)$.

$$|f(x) - f'(x)| \ \leq \ \frac{C_x \sigma^{\alpha(x)}}{\alpha(x)+1}$$

Analyse

$\Delta f(x) = |f(x) - f'(x)|$ **est la variation de fitness espérée dans le voisinage de** x**.**

Pour un $\sigma < 1$ **fixé, elle décroît quand** α **croît.**

Pour des rayons de mutation suffisamment petits, les fonctions plus lisses sont plus faciles à optimiser.

$\longrightarrow$ **Une mutation dépendant de la localisation** $\sigma = \sigma(x)$ **?**
$\longrightarrow$ **Réglée pour obtenir un majorant constant sur** $\Delta f(x)$ **tout le long de la trajectoire ?**

Une mutation dépendant de la localisation

$\frac{C_x \sigma^{\alpha(x)}}{\alpha(x)+1} = K$, **un paramètre défini par l'utilisateur.**

$$\sigma(x) = \left(\frac{K(\alpha(x)+1)}{C_x}\right)^{\frac{1}{\alpha(x)}}$$

Selon la valeur de $\frac{K}{C_x}$, le rayon de mutation peut

- **croître avec α (quand $\frac{K}{C_x} \leq 0.8$),**
- **ou décroître avec α (quand $\frac{K}{C_x} \geq 1$).**

Expériences avec N et U

Deux (1+1)ES ont été comparés :

- **avec mutation de rayon fixé : (ES),**
- **avec rayon de mutation adaptatif : (ESadapt).**

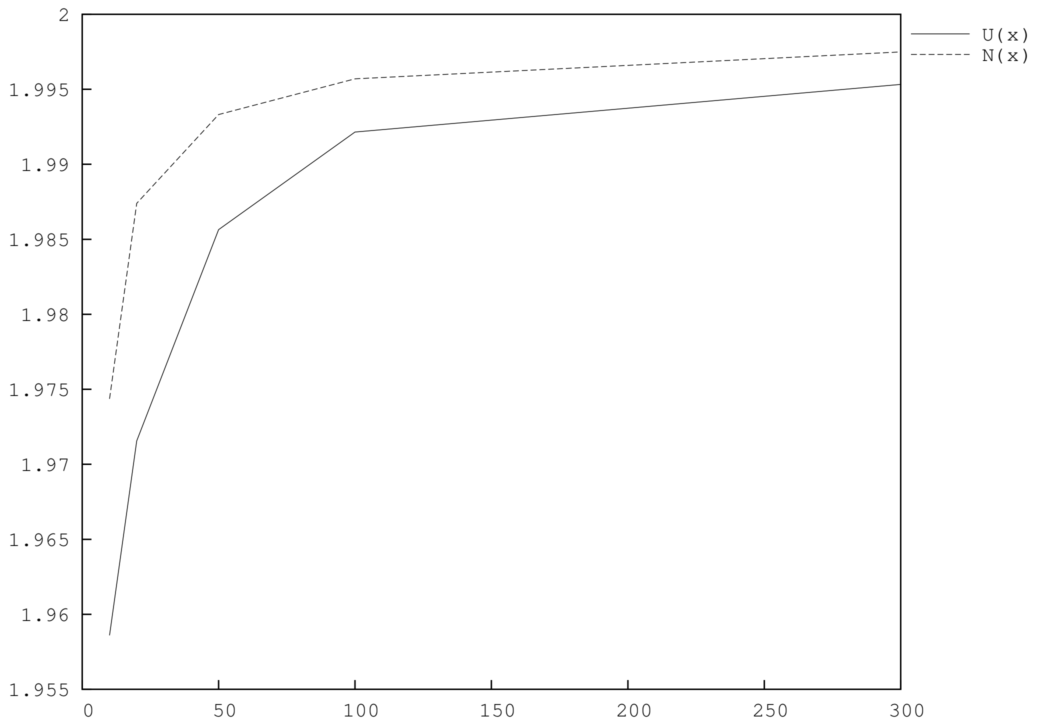
$$\sigma(x) = \beta \left(\frac{K(\alpha(x)+1)}{C_x} \right)^{\frac{1}{\alpha(x)}}$$

C_x **et** K **sont considérés comme constants sur** N **et** U**.**

β **varie de façon à avoir un rayon variable de valeur moyenne comparable à la valeur fixée** σ **de l'ES.**

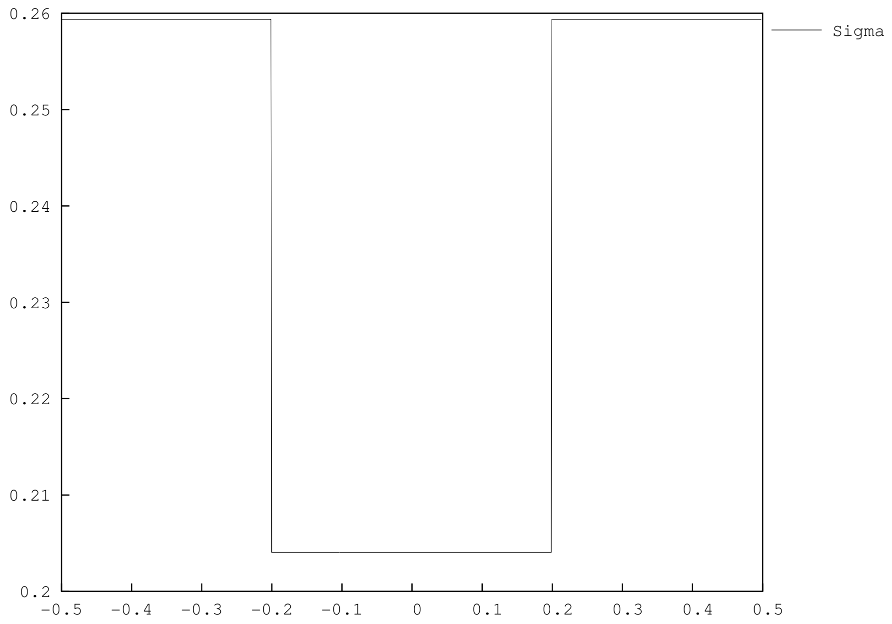
Les statistiques sont faites sur 100 runs pour chaque jeu de paramètres.

Recherche aléatoire pure

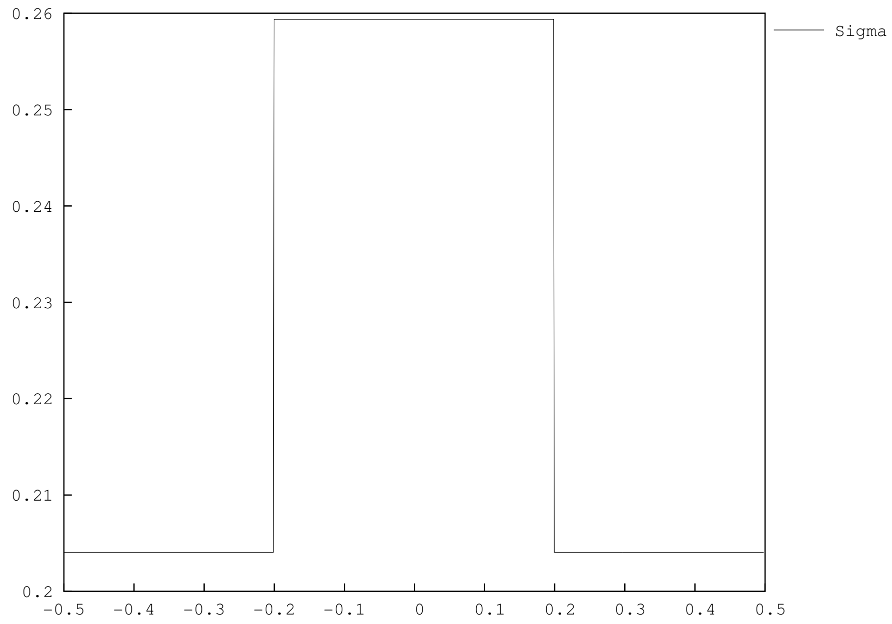


Résultats moyens (100 runs) d'un algorithme de recheche aléatoire pure sur U(x) et N(x), le nombre des évaluations est en abscisse.

Profils adaptifs $\sigma(x)$ pour N et U



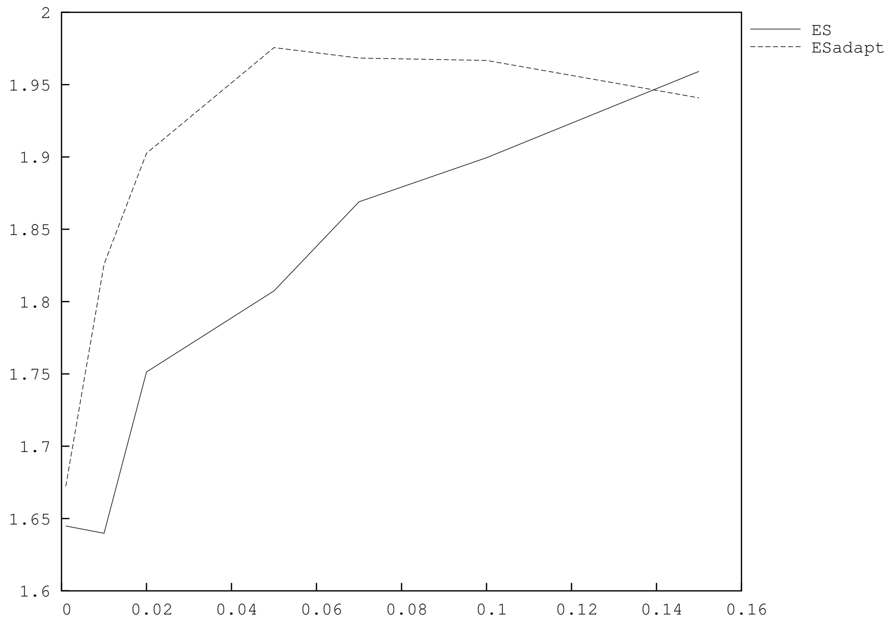
Pofil Sigma pour N(x)



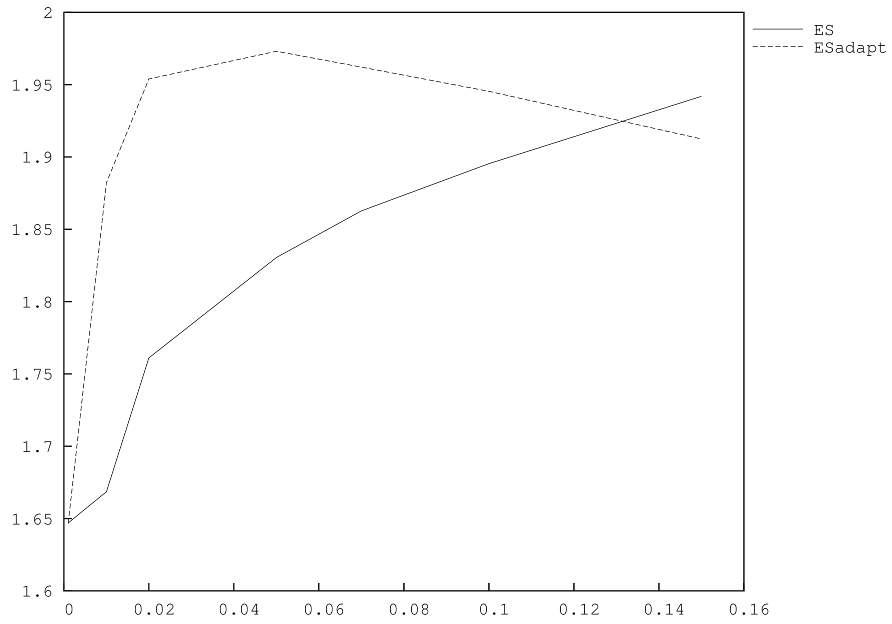
Pofil Sigma pour U(x).

$\longrightarrow$ **Un σ plus large pour les zones les plus irrégulères.**

10 générations d'un (1+1)ES

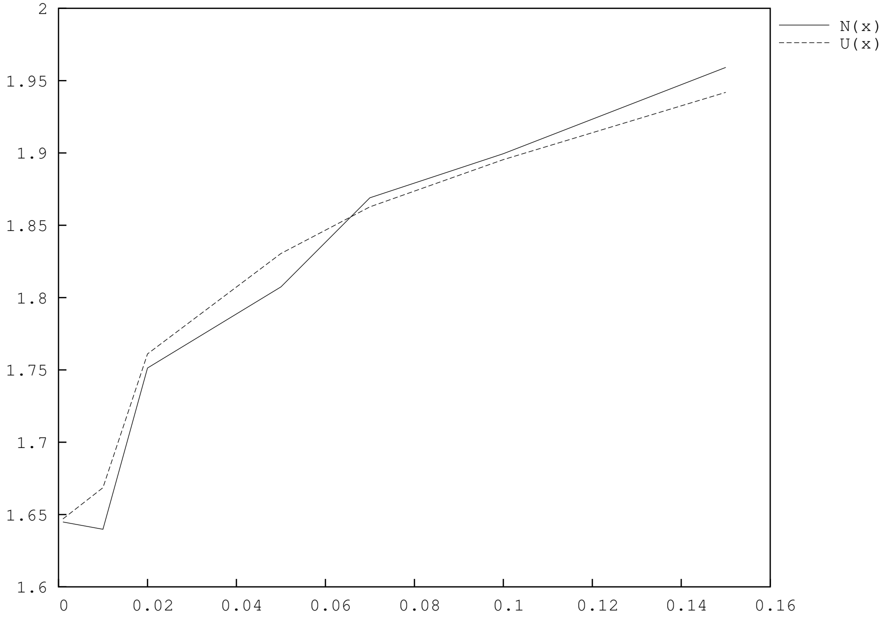


Résultats sur N(x).

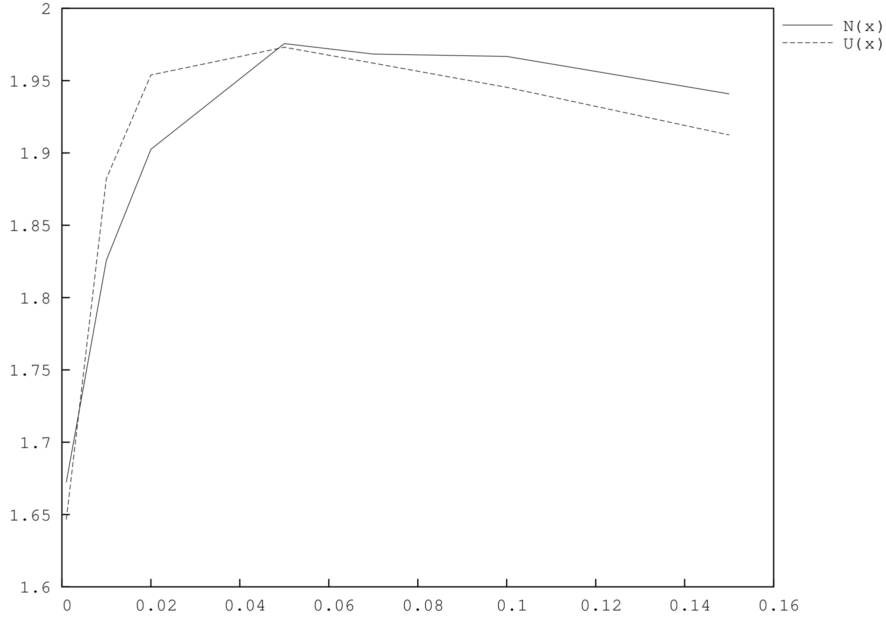


Résultats sur U(x).

Comparaison de U(x) et N(x), 10 générations



ES



ESadapt

Estimation des exposants en ligne

La méthode requiert le calcul de $\alpha(x)$ et de la norme Höldérienne C_x en chaque x :

- **:-(L'échantillonnage d'un voisinage a un coût calculatoire.**
- **:-) Les évaluations antérieures peuvent être utilisées.**
- **:-) Un AE échantillonne préférentiellement les régions intéréssantes (optima).**

Un routine d'estimation de C_x peut être intégrée à un $(1 + \lambda)ES$ avec très peu de de calculs additionnels.

Question actuelle : Design d'une routine efficace d'estimation de C_x et $\alpha(x)$ en ligne pour les $(\mu, \lambda)ES$ et les $(\mu + \lambda)ES$.

Estimation de C_x et de $\alpha(x)$

Echantillonnage de f sur un voisinage de taille ε autour de x : $f(x_i)$ pour $x_i = x-i/n, \ldots, x+i/n$ (en pratique, $i \simeq 3$).

Oscillation osc_ρ de f sur $B(x,\rho)$ pour $\rho = 1/n, \ldots, i/n$:

$$\mathbf{osc}_\rho = \sup_{y \in B(x,\rho)} f(y) - \inf_{y \in B(x,\rho)} f(y).$$

Régression aux moindres carrés du vecteur $(\log(\mathbf{osc}_\rho))_\rho$ selon $(\log(\rho))_\rho$:

- **$\alpha(x)$ est la pente,**
- **C_x est le point de croisement de la droite avec l'axe des ordonnées.**

Juliany & Vose “The genetic algorithm fractal”

Un modèle de système dynamique des AEs fondé sur un théorème des schémas avec égalité : les populations successives peuvent être représentées à l’aide d’une fonction $\mathcal{G}$.

A partir d’une population initiale aléatoire x, l’AE produit des populations successives : $\mathcal{G}(x)$, $\mathcal{G}^2(x)$, ..., $\mathcal{G}^n(x)$

⇒ Visualisation des bassins d’attraction $\mathcal{G}^\infty(x)$ de ce système dynamique produit des images fractales.

Juliany & Vose "The genetic algorithm fractal"