Quoi de neuf dans les algorithmes génétiques ?

Un bilan de 15 ans d'optimisation évolutionnaire

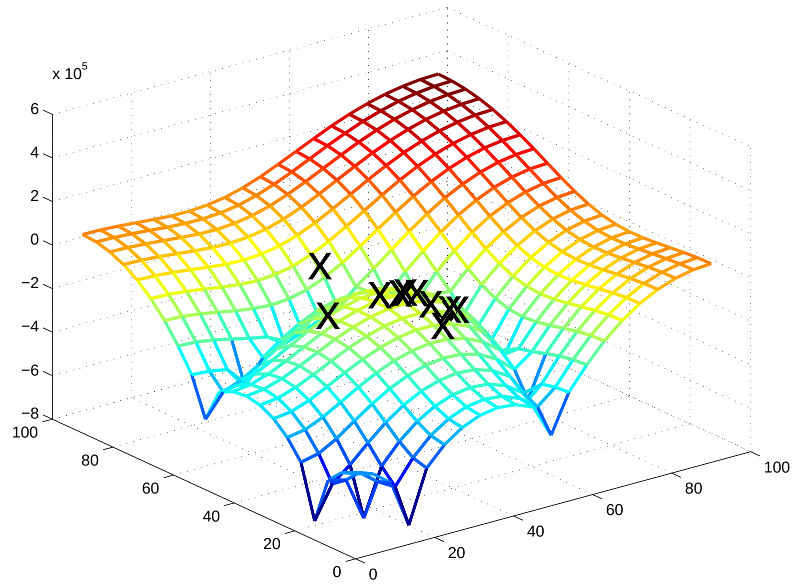
Rodolphe Le Riche, CNRS et Ecole des Mines de St Etienne

leriche@emse.fr

Le calcul évolutionnaire : définition

Des méthodes numériques basées sur des “populations” de points pour résoudre des problèmes complexes.

- Algos. génétiques (AGs)
- Strat. d’évolutions (ES)
- ...
- Beaucoup d’activité depuis 15 ans.



Algorithmes évolutionnaires (AE) pour l'optimisation

$$\begin{cases} \min_{x \in S} f(x) \\ S \equiv \Re^n \text{ ou } \mathcal{D}^n \text{ ou } \{\Re^{n1}, \mathcal{D}^{n2}\} \end{cases}$$

- Pas de condition particulière sur f ou S (pas rare en optimisation stochastique, cf. le recuit simulé, Monte Carlo ou les recherches taboues).

AEs : la métaphore Darwinienne

Les individus d'une espèce évoluent par reproduction et sélection pour maximiser leur performance dans leur environnement $\equiv$ résolution d'un pb. d'optimisation.

individu	x
chromosome (gènes)	codage de x
phénotype (caract. exprimés)	$f(x)$
population	$\{x^1, \ldots, x^\mu\}$

Cette métaphore facilite l'explication, mais ne justifie pas les choix algorithmiques !

Structure d'un AE

$t \leftarrow 0$

`Initialiser la pop.`

`Evaluer la pop.` (f)

Tant que continuer

$t \leftarrow t+1$

`Sélection.`

`Reproduction (croisement, mutation).`

`Evaluer les enfants.`

`Remplacer certains parents par les enfants.`

Fin.

Plan de la présentation

90 - 95	l’essort des AGs, le rêve d’un AE universel
95	NFL et la fin de l’AE universel
95 - auj.	la spécialisation
auj.	quelques tendances

Période 90-95

L’essort des AGs, le rêve d’un AE universel

l'AE dominant des années 90 : l'algorithme génétique

J. Holland (1975), D. Goldberg (1989)

Représentation binaire, x est écrit $[0\ 1\ 1 \dots 0]$

Croisement $\begin{array}{l}[0|1\ \ 1\ \ 1] \\ [1|1\ \ 0\ \ 0]\end{array} \rightarrow [0\ 1\ 0\ 0]$

Mutation $[0\ 1\ 0\ 0] \rightarrow [0\ 1\ 0\ 1]$

Sélection $\forall\ x^i$ et x^j dans population,
$f(x^i) \leq f(x^j) \ \Rightarrow \ P_{\text{sél}}(x^i) \geq P_{\text{sél}}(x^j)$

Théorie des AGs : les schémas

Schéma $\equiv$ sous-ensemble de $\mathcal{S}$, $\mathcal{H} = \{0 \ * \ 1 \ * \ 0\}$.

L'essort des AGs :

- Les schémas courts et performants (BBs) se propagent dans la population par sélection/croisement.
- Parallélisme implicite : un bon individu favorise tous les schémas auxquels il appartient ($\Rightarrow$ alphabet binaire).

$\Rightarrow$ espoirs d'algorithmes performants en moyenne au début des années 90.

Mais :

- Estimation (biaisée) de $\bar{f}(\mathcal{H})$.
- Les bons schémas ne contiennent pas nécessairement l'optimum (problèmes difficiles sont non-linéaires).

Les grands débats des années 90 (I)

- La représentation

codages binaires contre	codages naturels
Prc : faire automatiquement émerger les BBs	Prc : il existe une paramétrisation naturelle (on connait les BBs, relations d'équivalences).
Codage binaire Gray, évolution du codage (inversion, Holland 75; Messy GAs, Goldberg 91 à 00), . . .	Vecteurs de nombres réels (Stratégies d'évolution), cellules de Voronoï en optim. topologique (Schoenauer 94 à 00), . . .

Les grands débats des années 90 (II)

- Ce qui fait avancer un AE c'est ...

le croisement	**vs. la mutation**
le mélange des BBs	les perturbations + la sélection
AGs (Goldberg)	Stratégies d'évolution (Schwefel, Bäck), Programmation évolutionnaire (L. et D. Fogel)

- Le réglage de paramètres : taille de population, pression de sélection, probabilités de mutation et de croisement.

Vers une spécialisation des AEs

- Années 90 : des résultats théoriques et empiriques contradictoires (codages, algos, paramètres).
- Progressivement, généralisation de l'idée de **spécialisation** des AEs au problème ($\neq$ algo. universel).
- Un résultat théorique vient conforter cette tendance : le théorème du "No Free Lunch".

Le théorème du No Free Lunch

En moyenne sur tous les problèmes d'optimisation, le comportement de n'importe quel algorithme est le même.
Wolpert et Macready, 1995

- Interprétation : ce qu'un algorithme gagne sur certains problèmes est perdu sur d'autres.
- En pratique, on ne considère pas tous les problèmes, on exige au moins une certaine régularité de f sans laquelle le problème de l'optimisation globale ne peut pas être résolu.

Période 95-auj.

La spécialisation

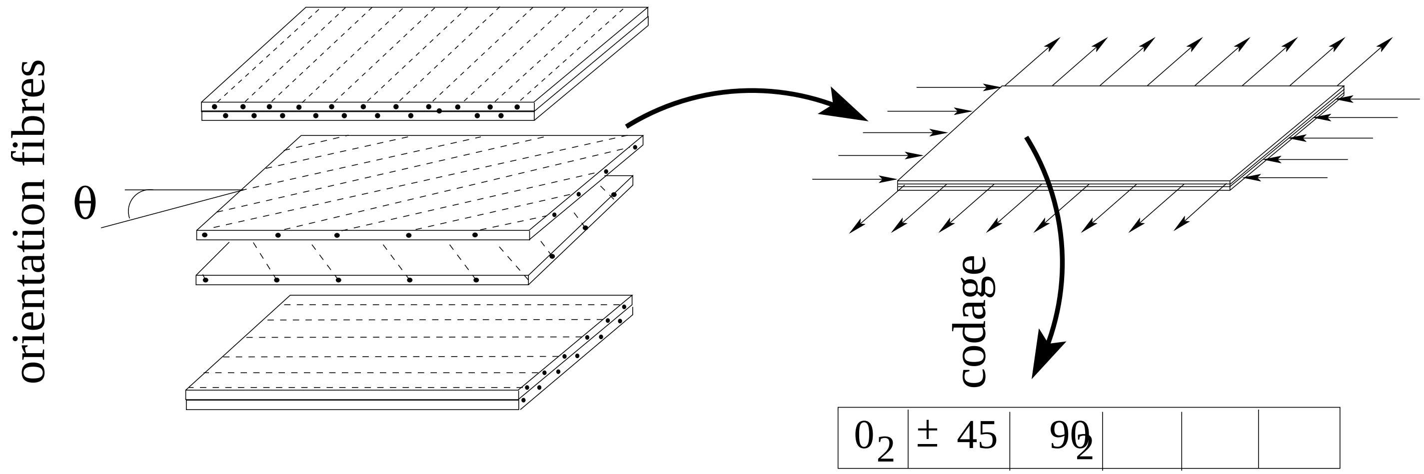
Expl. de spécialisation des AEs :
AEs comme méta-heuristiques

- Les AEs gagnent à être couplés à des méthodes d'optimisation locales ou à d'autres heuristiques ("adaptation vs. apprentissage").
- Couplages en parallèle ou en série :

```
t ← 0, initialiser la pop.
Evaluer la pop.  (f)
Tant que continuer
    t ← t+1
    Sélection.
    Reproduction (croisement, mutation, heuristique).
    Evaluer les enfants.
    Remplacer certains parents par les enfants.
Fin tant que .
heuristique
```

Expl. de couplage avec une heuristique : mise à l'échelle (1)

(Optimisation de stratifiés composites, R. Le Riche 94)



$$\begin{cases} \min_{N,\theta_i} N & \text{(ou épaisseur } h = Nh_i\text{)} \\ \lambda_{\text{flamb}} \geq 1 & \text{rupture flambement} \\ \lambda_\varepsilon \geq 1 & \text{rupture déformations principales} \end{cases}$$

Mise à l'échelle (2)

Connaissance RDM pour estimer une nouvelle épaisseur de plaque :

$$\lambda_\varepsilon \approx h \quad , \quad \lambda_{\text{flamb}} \approx h^3.$$

$$\begin{cases} h_{\text{flamb}} = h/\sqrt[3]{\lambda_{\text{flamb}}} \\ h_\varepsilon = h/\lambda_\varepsilon \end{cases} \Rightarrow \quad h = \text{arrondi}_{Nh_i}\left[\max(h_{\text{flamb}}, h_\varepsilon)\right].$$

- Appliqué sur 10% des cas.
- La mutation peut aussi changer l'épaisseur (opt. globale).
- L'algorithme gagne 10% d'efficacité (80% de chances de trouver un optimum pratique en 1310 analyses parmi plus de 10 millions de possibilités).

Période 00-auj.

Deux voies prometteuses

Tendances récentes (I) :
algos. à estimation de densités (EDAs)

- Les EAs définissent implicitement (à travers le croisement et la mutation) une densité de probabilité d'échantillonner un nouveau point, $P(x)$.
- Idée des EDAs (Baluja 94, Mülhenbein 96) : expliciter $P(x)$ qui remplace les opérateurs génétiques $\Rightarrow$ meilleure formalisation (Bayes, Markov)

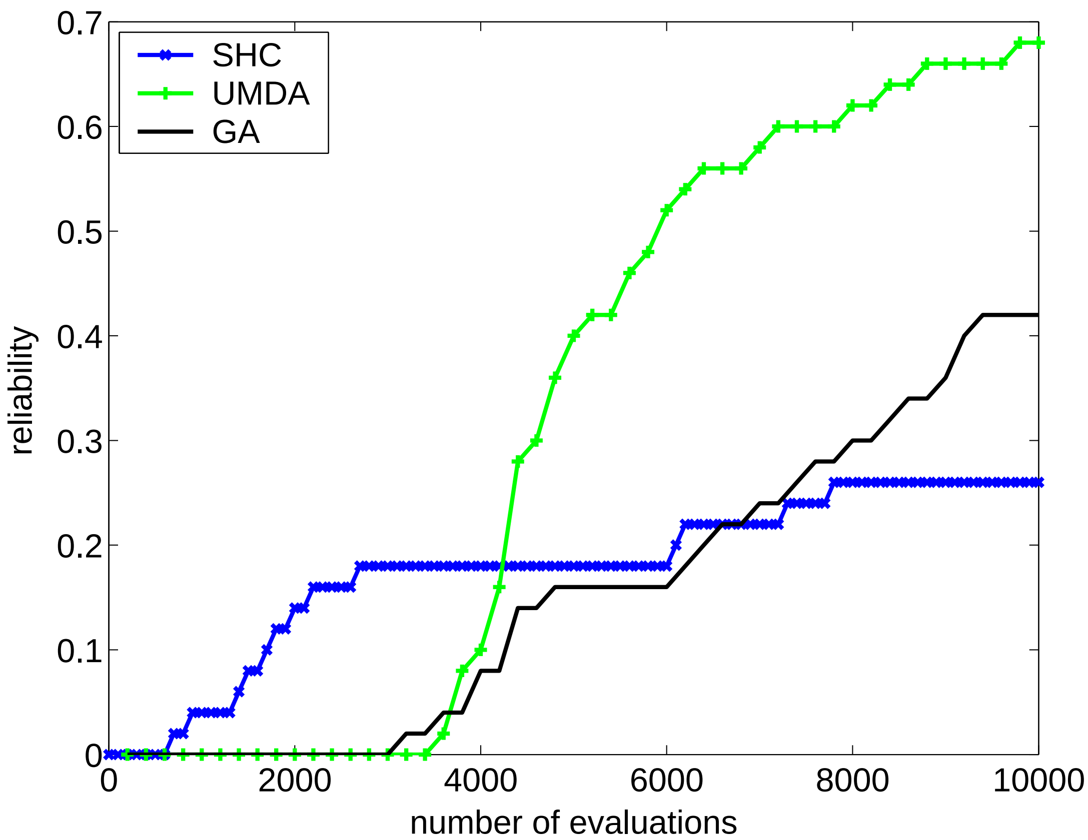
$$P^{t+1}(x) = \frac{P^t_{\text{sélection}}(x).P^t(x)}{\text{normalisation}}$$

- Difficultés : le choix de P est un compromis entre sa précision et sa stabilité.

EDAs : exemple en optimisation de composites

(Grosset et al. 2004)

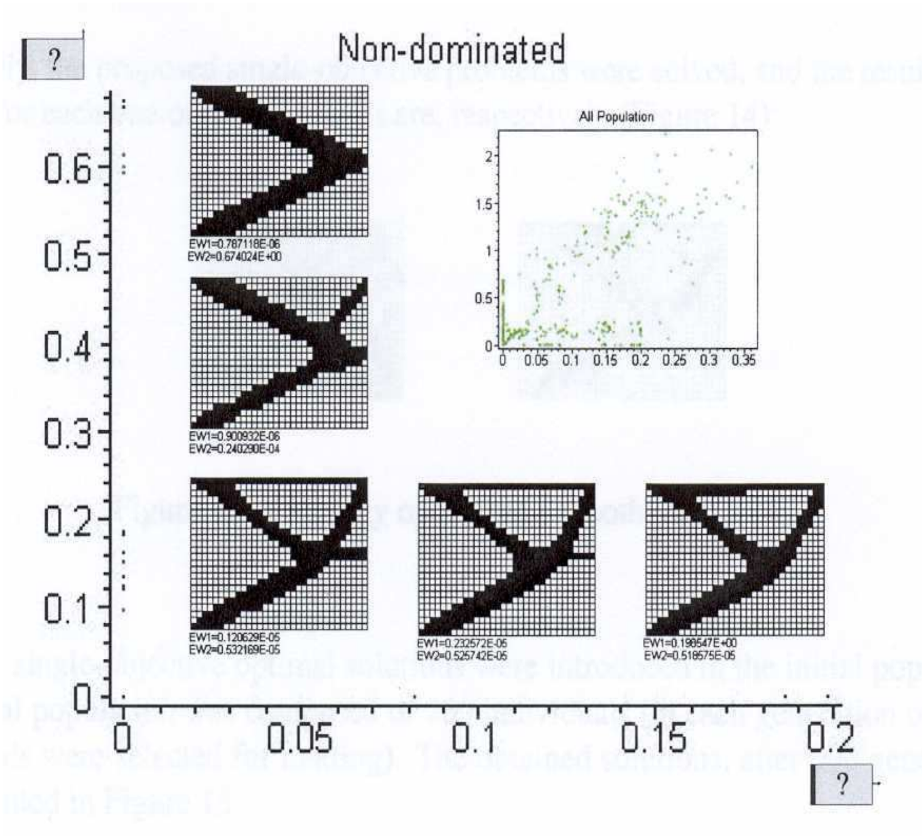
$$\begin{array}{ll} \textbf{max} & f_1(\theta_1, \ldots, \theta_{15}) \\ \textbf{t.q.} & \nu_\mathrm{l} \leq \nu_\mathrm{eff} \leq \nu_\mathrm{u} \end{array}$$



Tendances récentes (II) : optimisation multi-critères

(J.F. Aguilar Madeira, 2002)

$$\begin{cases} \min_{x \in S} f_1(x) \\ \dots \\ \min_{x \in S} f_m(x) \end{cases}$$



Ensemble des solutions = ens. des meilleurs compromis = front de Pareto

Les AEs bénéficient de leur population.

Conclusions

- Des méthodes aussi populaires que critiquées depuis 15 ans.
- Les AEs vont (contribuer à) renouveller l'optimisation en déplaçant les centres d'intérêts de l'efficacité vers la représentation, l'extraction de connaissances d'une population, la collaboration entre méthodes ...
- Quel algorithme pour quel problème ? (utilisation de la corrélation f-distance, $\ldots$)

Backup slides

Calcul évolutionnaire

- Les algorithmes génétiques : J. Holland 75 (psychologie, biologie), D. Goldberg 89 (SPI).
- Les stratégies d'évolution : I. Rechenberg 65, H.-P. Schwefel 81, T. Bäck 95 (optimisation).
- La programmation évolutionnaire : L.J. Fogel (62) (prog. d'automates), D.B. Fogel (88).
- La programmation génétique : J. Koza (94) (programmation automatique).
- L'optimisation statistique (EDAs, Bayesienne) : Baluja (94), Mühlenbein (99).

$\Longrightarrow$ 30 ans d'histoire, des milliers d'applications dans tous les domaines, > 10 conférences intl. par an, > 3 revues.

La spécialisation des AEs (I) : l'auto-adaptation

Les paramètres de l'AE sont sur le chromosome

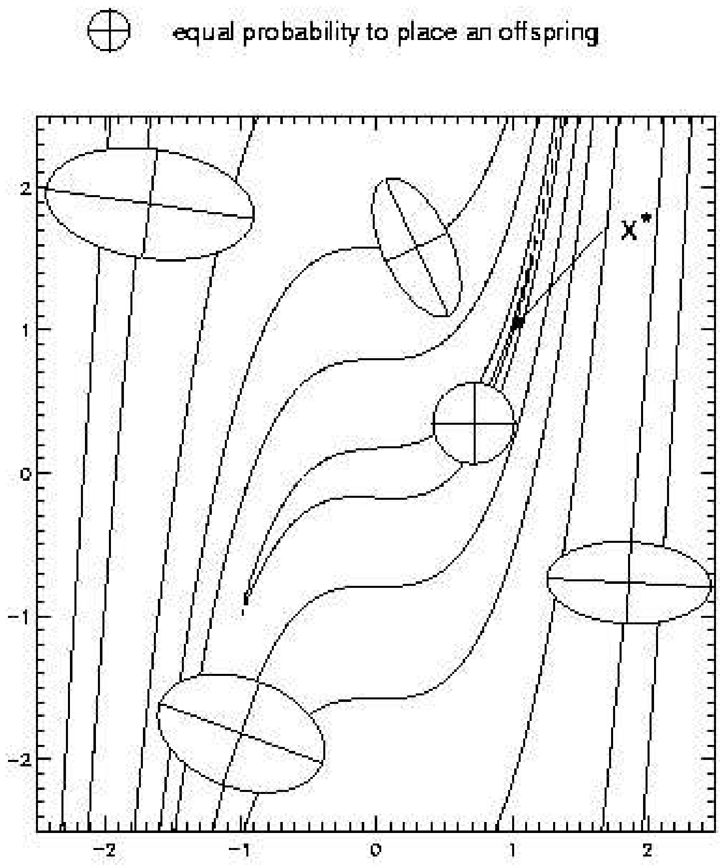
$$[\quad x \quad prob\_mutation \quad prob\_croisement \ldots \quad]$$

Ils subissent les opérations génétiques et sont donc adaptés comme x.

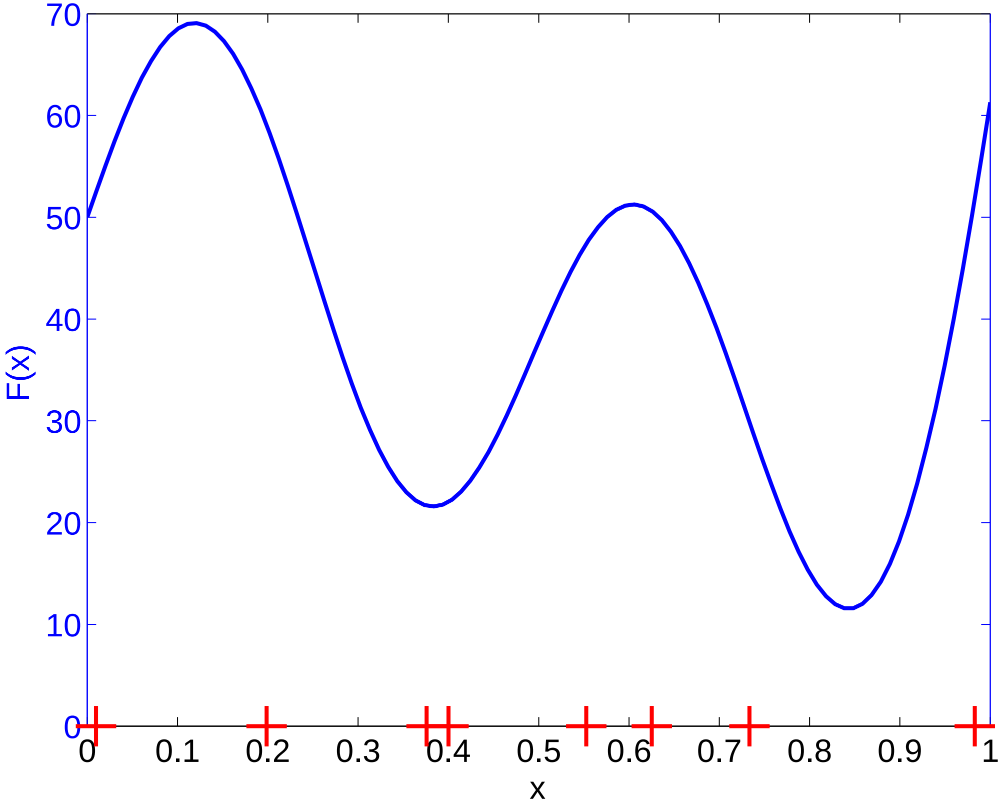
L’auto-adaptation (exemple)

Auto-adaptation de la mutation dans ES (Schwefel 77, Bäck 91).
Mutation = perturbation Gaussienne, $x' \; = \; x + \mathcal{N}(O, C)$.

1. Muter la mutation,
   $C \; \rightarrow \; C'$
2. Utiliser C’,
   $x' \; = \; x + \mathcal{N}(O, C')$
3. Evaluer $f(x')$,
   sélectionner $[x' \quad C']$.

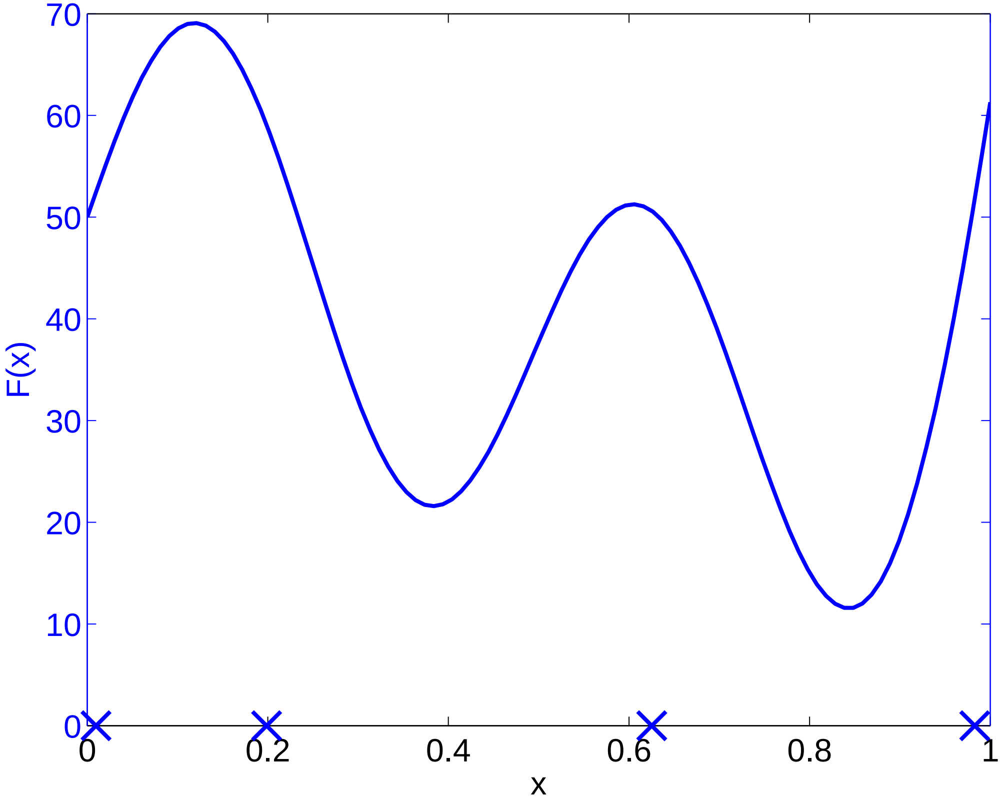


EAs : estimation de densité implicite



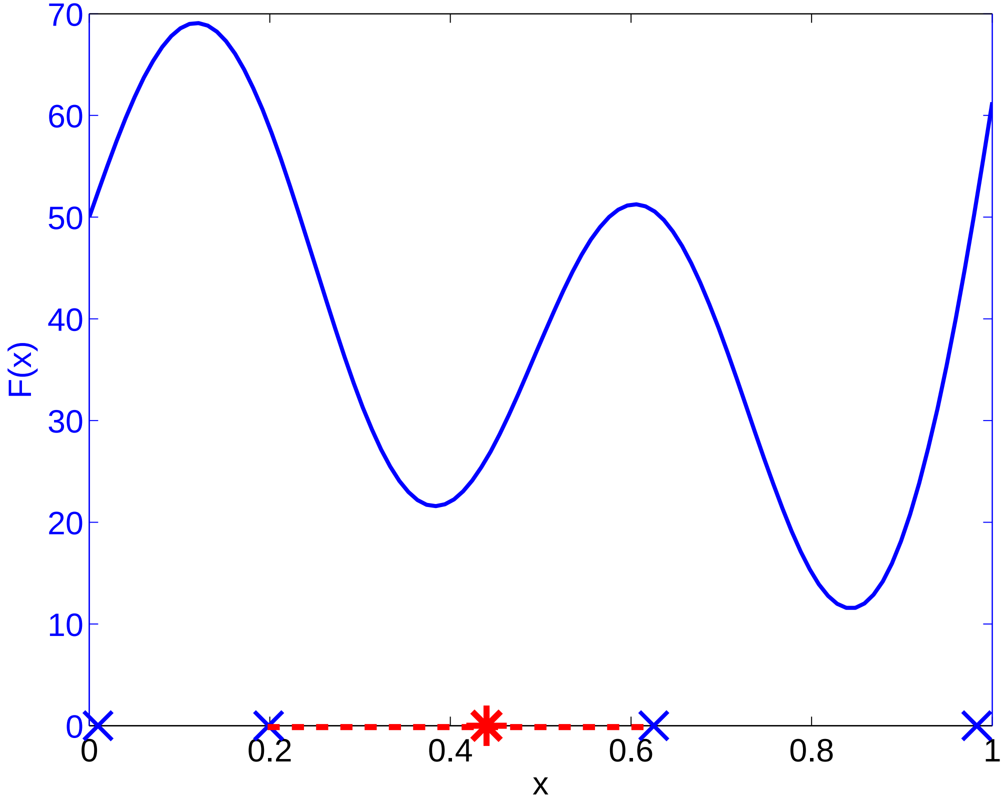
population initiale uniforme

EAs : estimation de densité implicite



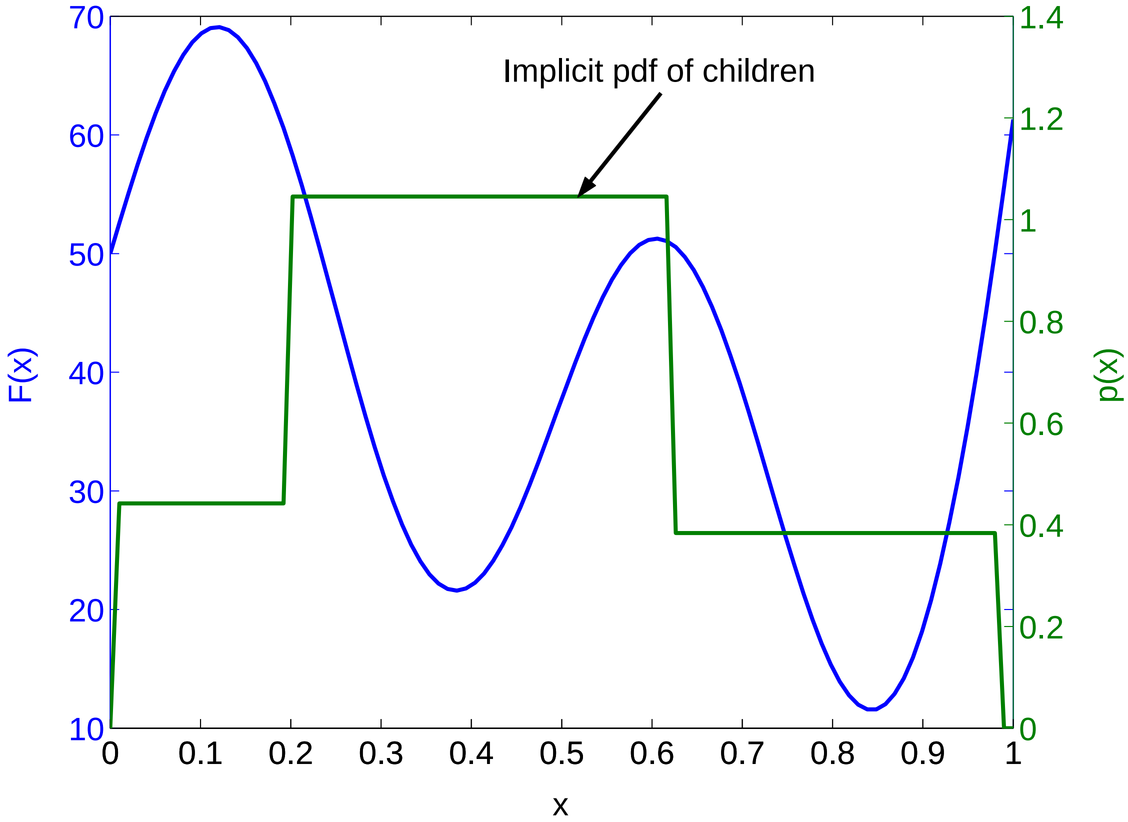
points gardés par sélection

EAs : estimation de densité implicite



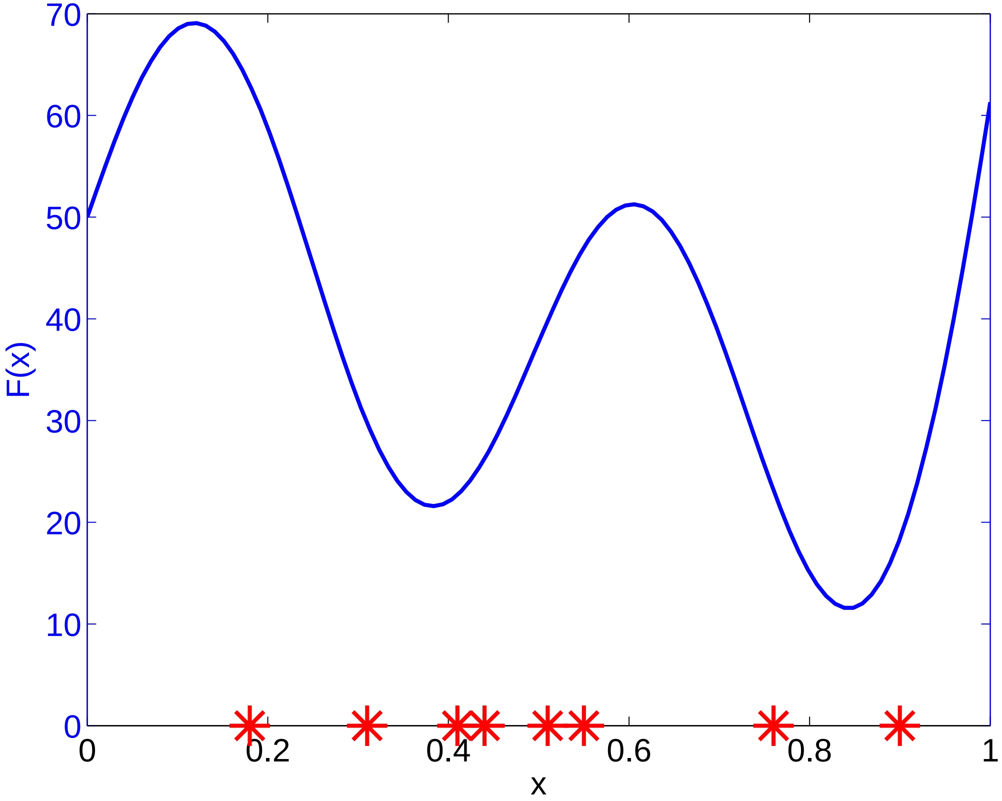
création de nouveaux points par croisement
(weighted average in $\mathbb{R}^n$)

EAs : estimation de densité implicite



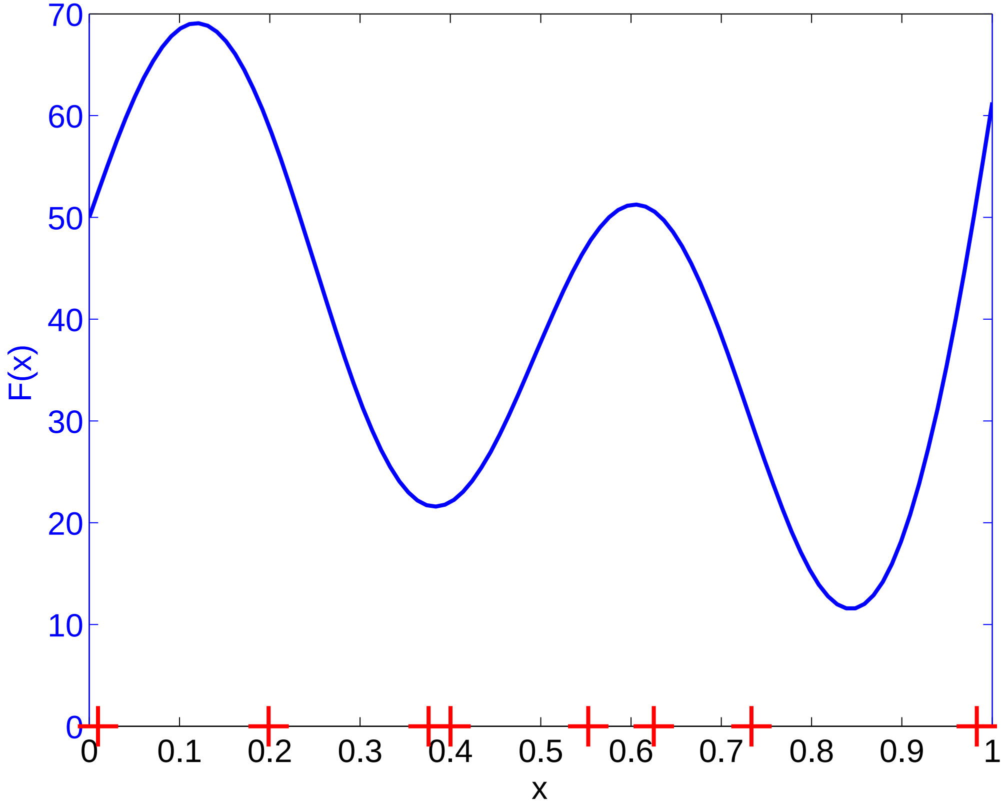
création de nouveaux points par croisement

EAs : estimation de densité implicite



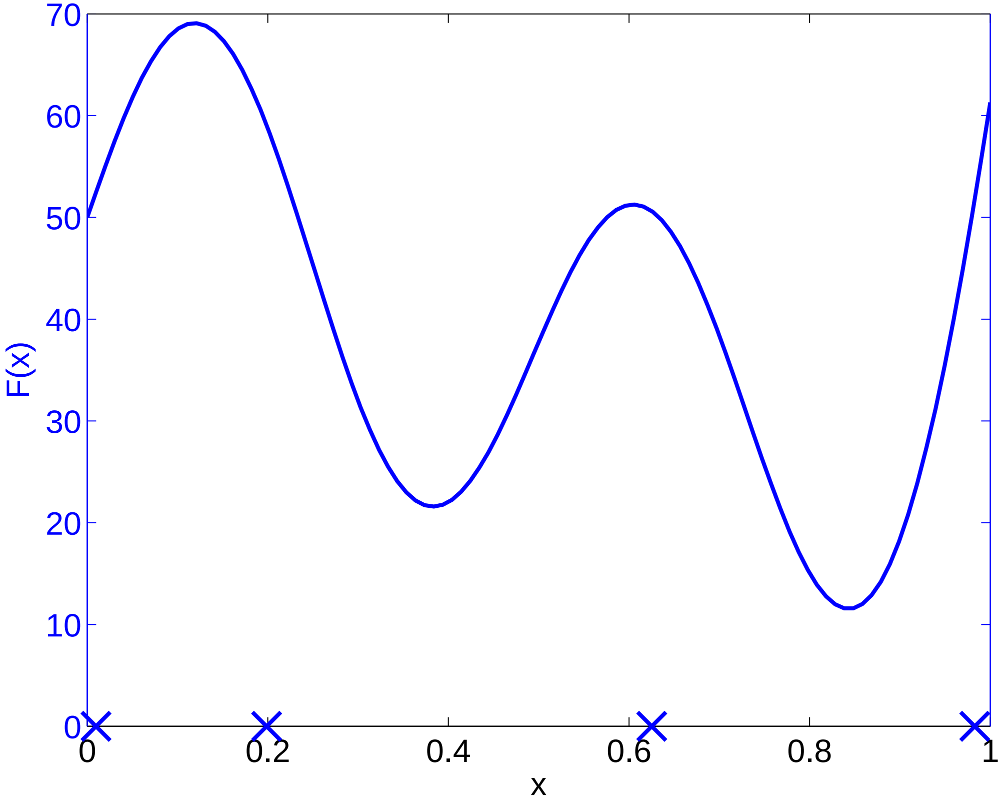
nouvelle population

Tendances récentes (I) : fonctionnement schématique d'un EDA



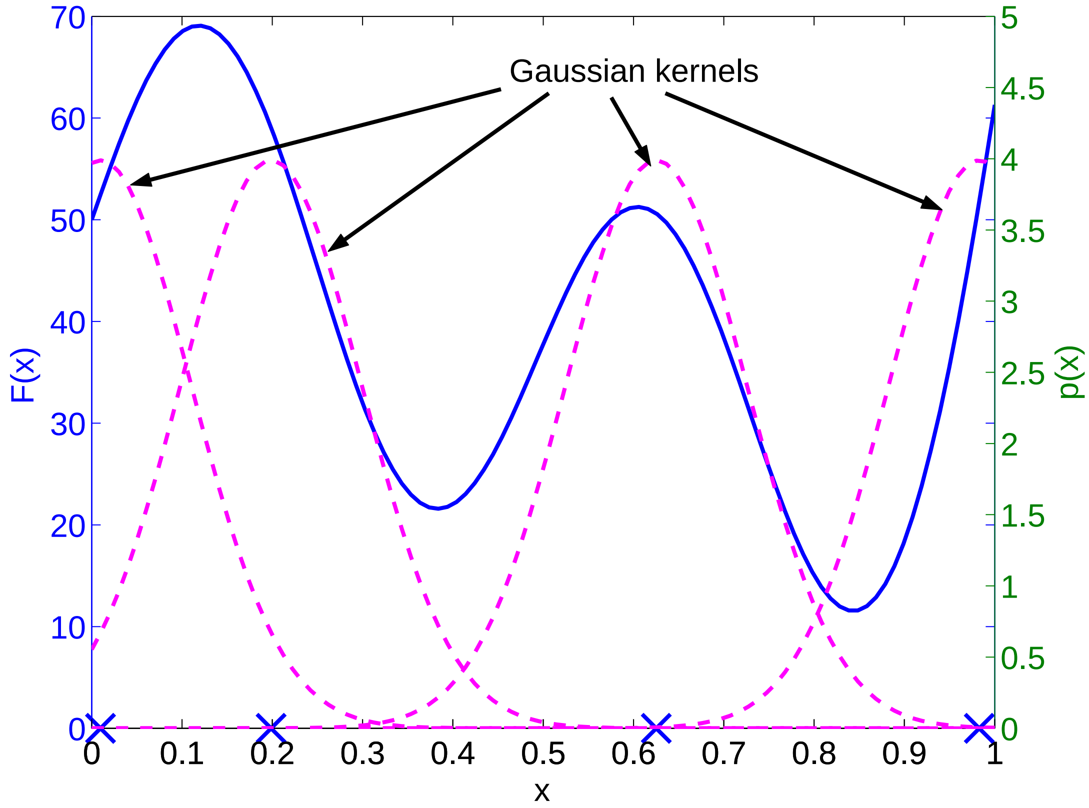
population initiale uniforme

Tendances récentes (I) :
fonctionnement schématique d'un EDA



points gardés par la sélection

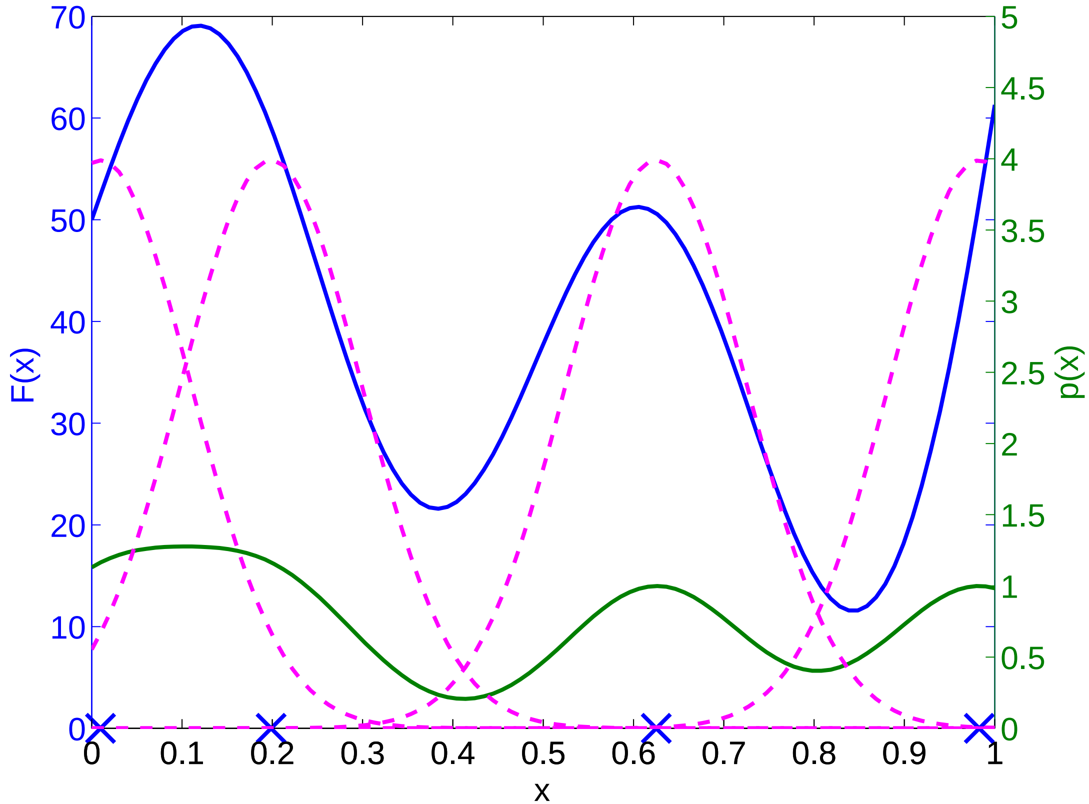
Tendances récentes (I) : fonctionnement schématique d'un EDA



estimation de la densité des bons points $P(x)$

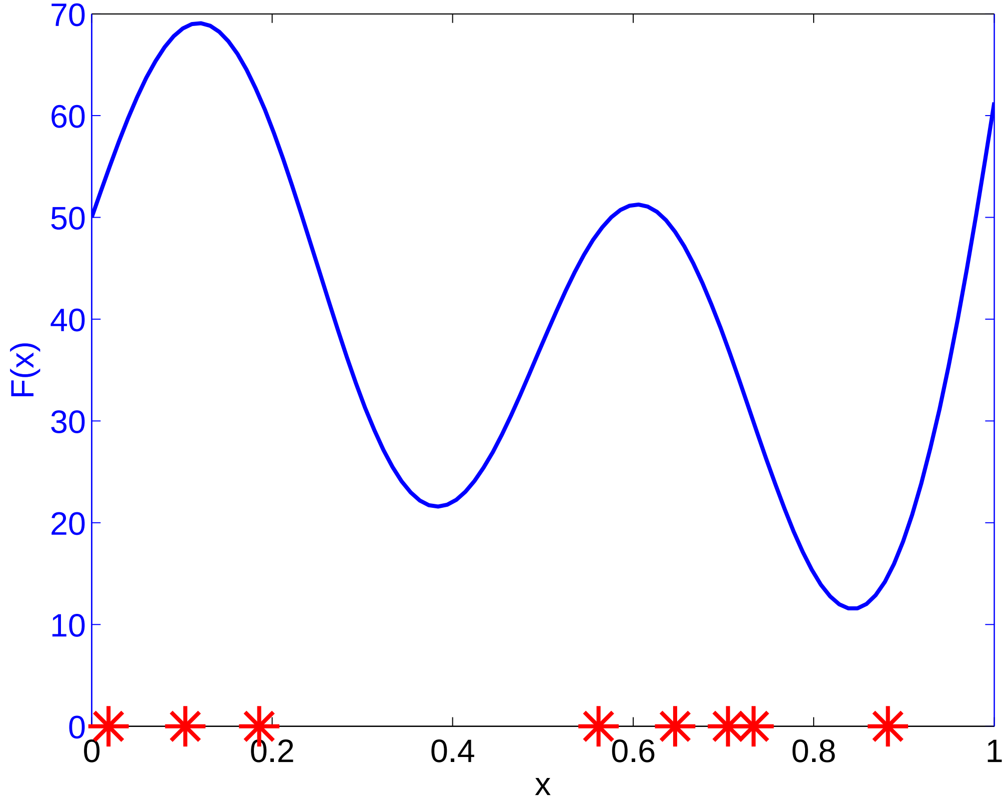
(noyaux, $P(x) =$ const$/N \sum_{i=1}^{N} \exp(-(x - x_i)^2/(2\sigma^2))$)

Tendances récentes (I) : fonctionnement schématique d'un EDA



densité des bons points estimée $P(x)$

Tendances récentes (I) : fonctionnement schématique d'un EDA



nouvelle population obtenue par échantillonnage de $P(x)$

Tendances récentes (II) : optimisation multi-critères

$$\begin{cases} \min_{x\in S} f_1(x) \\ \dots \\ \min_{x\in S} f_m(x) \end{cases}$$

Exemple

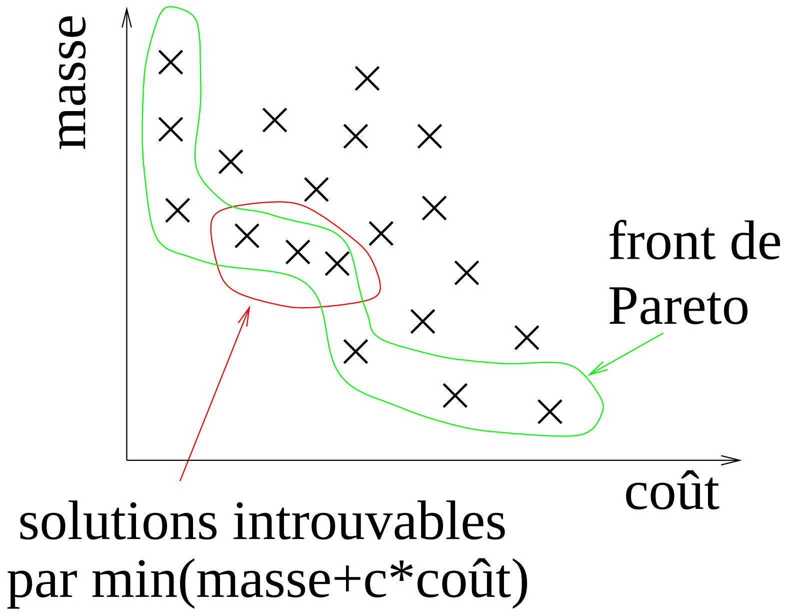
$$\begin{cases} \min_{x\in S} \text{masse}(x) \\ \min_{x\in S} \text{coût}(x) \end{cases}$$

Ensemble des solutions = front de Pareto = ens. des points non dominés.

x domine y ssi,

$\forall i, f_i(x) \leq f_i(y)$ et

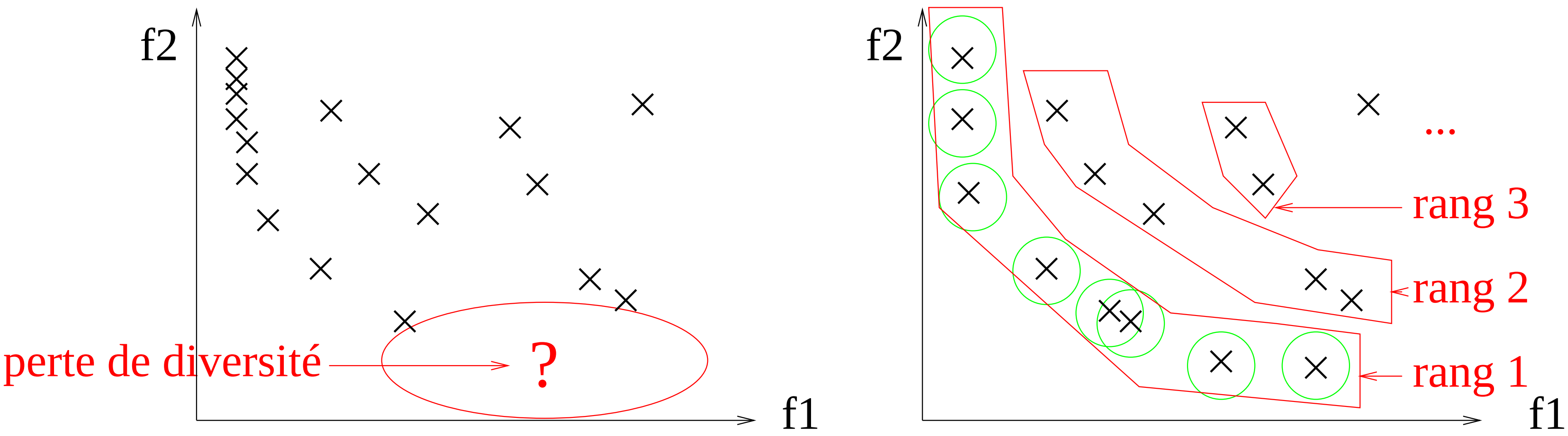
$\exists j / f_j(x) < f_j(y)$



Tendances récentes (II) : optimisation multi-critères par AEs

(J.D. Schaffer 85, J. Horn et al. 94, K. Deb 98)

- Essentiellement le calcul de la performance est modifié / AE monocritère.
- Calcul de la performance :
  - Utilise la domination de Pareto.
  - Préserve la diversité (dans esp. des x ou des f).

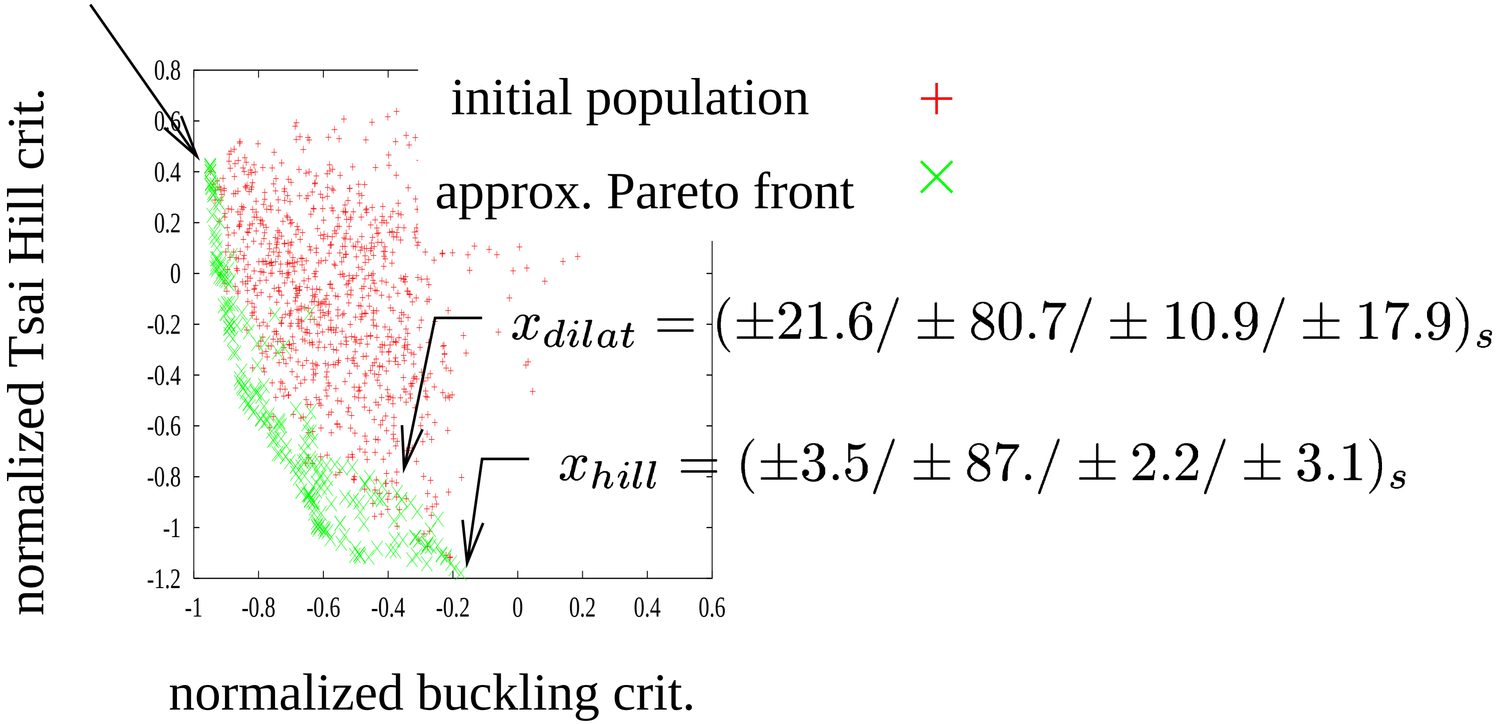


- L'élitisme a besoin d'une archive.

Optimisation multicritères de composites

(R. Le Riche, 2001)

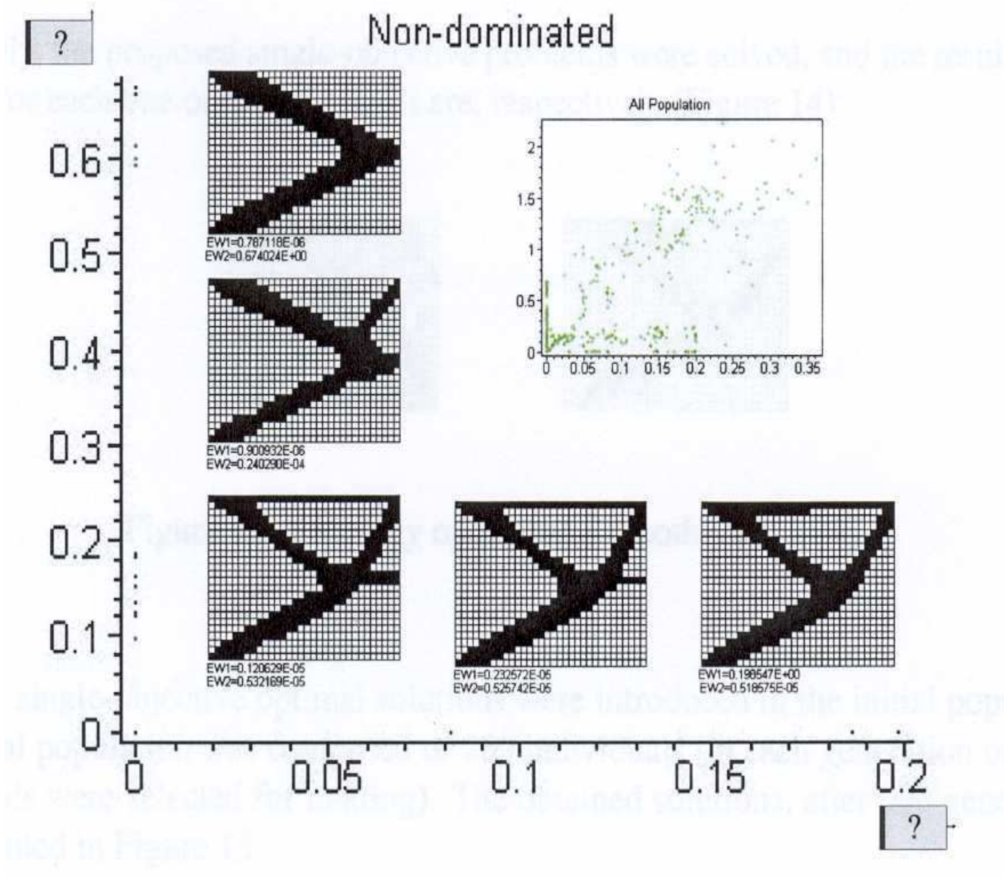
$$x_{buck} = (\pm 46.3 / \pm 44.8 / \pm 46.6 / \pm 43.)_s$$



Projection du front de Pareto des critères de dilatation thermique longit., flambement, et rupture (Tsai-Hill) dans le plan (Tsai-Hill , flambement). Graphite / epoxy, $1 \times 1\ m$, $N_x = -100000.\ N$, $N_y = 10000.\ N$. Niched Pareto GA.

Optimisation multicritères de topologies

(J.F. Aguilar Madeira, 2002)



Struct. de volume constant. Non Dominated Sorting GA + sélection avec clustering.