



# Techniques avancées

Marc Schoenauer

Equipe TAO – INRIA Futurs et LRI

<http://tao.lri.fr/>

Cours **Évolution Artificielle et Applications à la Robotique**,  
le 4 janvier 2010



## Plan

- Techniques de nichage
- Prise en compte des contraintes
- Optimisation multi-objectif
- Parallélisation



## Techniques de nichage

**Objectif** : identifier plusieurs optima (quasi) globaux

**Techniques** :

- Voisinage **explicites** i.e. le partage
- Voisinage **implicite** pas de paramètre de voisinage à fixer

Historiquement, le **partage** (1987), encore le plus utilisé



## Partage

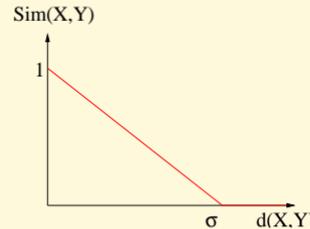
### Goldberg-Richardson 87

**Buts** : Préserver la diversité génétique

- pour éviter la convergence prématurée ... vers un optimum local
- pour permettre l'identification de plusieurs optima locaux  
ou quasi-optima

**Moyen** : Pénaliser les regroupements

- $Sim(X, Y) = \left(1 - \frac{d(X, Y)}{\sigma}\right)^+$



- Sélection **proportionnelle** basée sur  $\mathcal{F}'(X) = \frac{\mathcal{F}(X)}{\sum_{Y \in \mathcal{G}} Sim(X, Y)}$



## Partage Exemple

- 1

- 3

- 2

$$\begin{aligned}d(1, 2) > \sigma &\implies \text{Sim}(1, 2) = \text{Sim}(2, 3) = 0 \\d(1, 3) \ll \sigma &\implies \text{Sim}(1, 3) \approx 1\end{aligned}$$

donc

$$\left\{ \begin{array}{l} \mathcal{F}'(1) = \frac{\mathcal{F}(1)}{2} \\ \mathcal{F}'(2) = \mathcal{F}(2) \\ \mathcal{F}'(3) = \frac{\mathcal{F}(3)}{2} \end{array} \right.$$



# Optimisation d'interféromètres

L. Taieb – CMAP – 1997

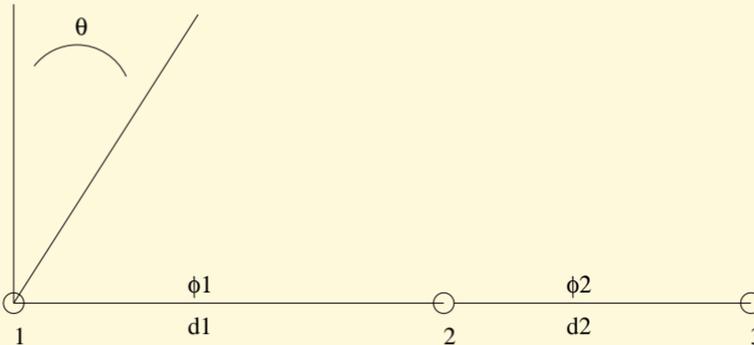
- Un **vrai** problème
- Solutions calculées pour 3 antennes, supposées pour 4 antennes.
- Pas de solution connue pour plus de 4 antennes.
- Pas de formulation analytique de la fonctionnelle à optimiser, mais un calcul numérique



## Le problème

Trouver la direction d'un signal électromagnétique reçu par un système de  $n$  antennes à partir des différences de phases mesurées entre antennes.

**Exemple** : Interféromètre à trois antennes alignées



$$\phi_{im} = \frac{2\pi d_i}{\lambda} \sin \theta + 2k_i\pi$$



## Le problème (2)

**Difficulté** Les déphasages sont connus à  $2\pi$  près, et avec des erreurs. Pour de petites erreurs de mesures, le déphasage théorique peut être déduit du déphasage mesuré ( $-\pi \leq \phi_{im} \leq \pi$ ) si les  $k_i$  sont connus.

$$\phi_{im} = \frac{2\pi d_i}{\lambda} \sin \theta + 2k_i\pi \pm \delta\phi_i$$

**But** Maximiser l'**erreur tolérable**  $\delta\phi$  permettant de retrouver sans ambiguïté les  $k_i$ , et donc  $\theta$ .

### Variables

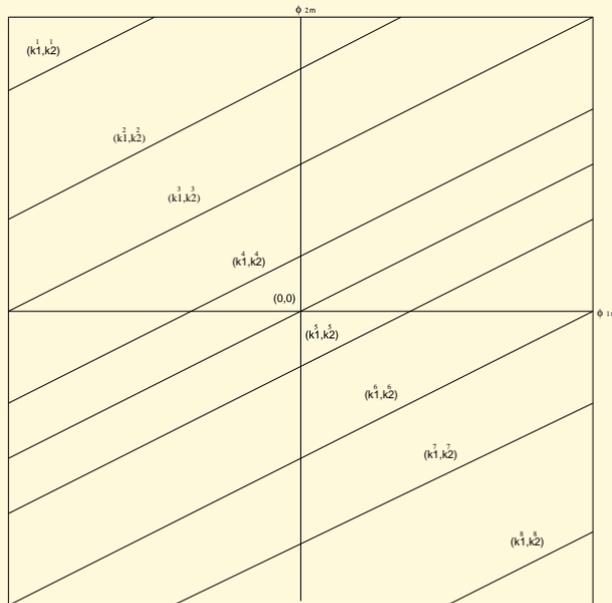
Le nombre d'antennes

La distance entre les antennes



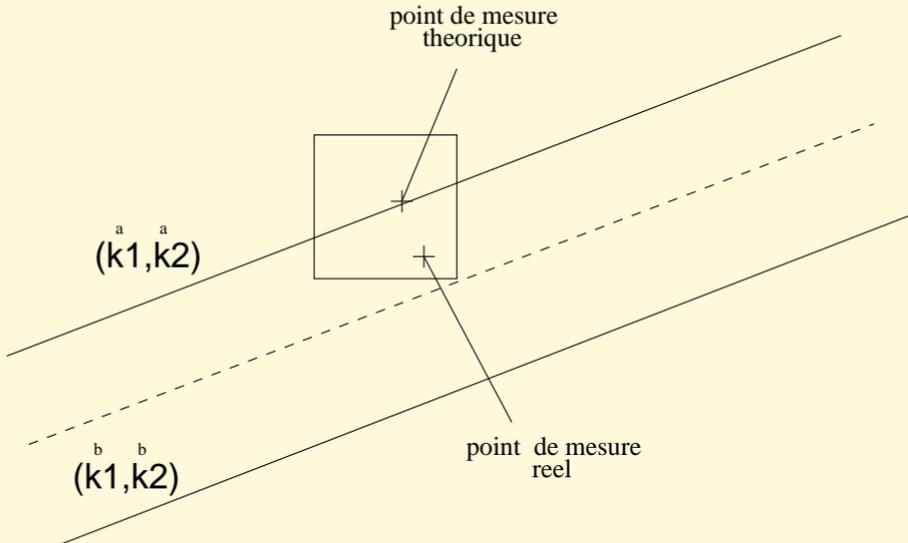
## Sans erreurs de mesure

$$\phi_{2m} = \frac{d_2}{d_1} \phi_{1m} + 2\pi \left( \frac{d_2}{d_1} k_1^a - k_2^a \right)$$





## Le cube erreur...

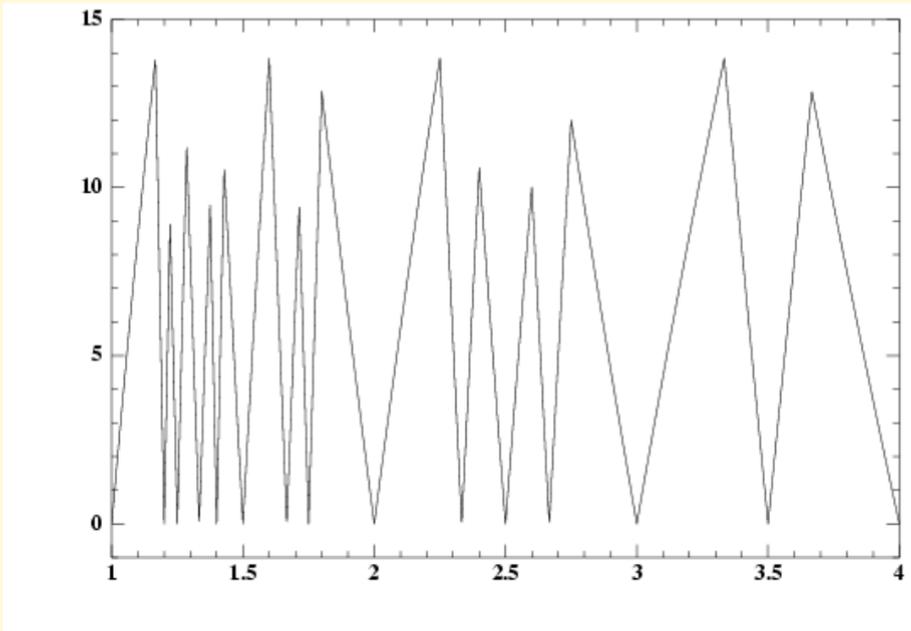


... ne doit pas recouvrir la droite mediane entre deux droites théoriques

→  $\delta\phi_{max}$



## Une fonctionnelle très chahutée



$$\delta\phi = f\left(\frac{d_1}{d_2}\right)$$



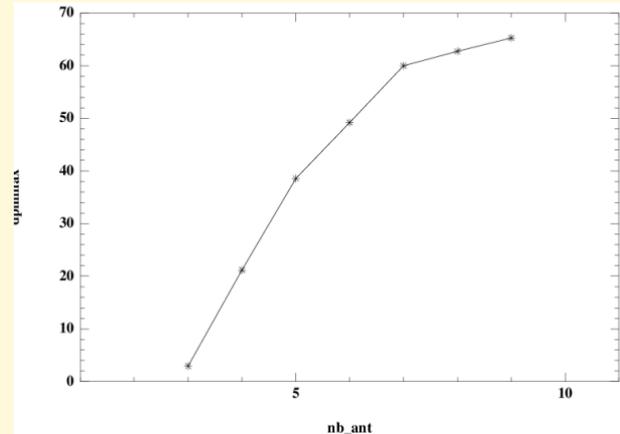
## AGs utilisés 1995 !

- Codage réel
- Sélection proportionnelle (roulette) –  $P_S$  1.2
- Remplacement générationnel
- Croisement: barycentrique ou par échanges de variables
- Mutation: Ajout de bruit gaussien d'écart-type fixe



## Résultats

- Optimisation de configurations jusqu'à 10 antennes.
- Configurations pour 4 antennes quelquefois meilleures que les solutions connues.
- Calcul de **tous** les optima.



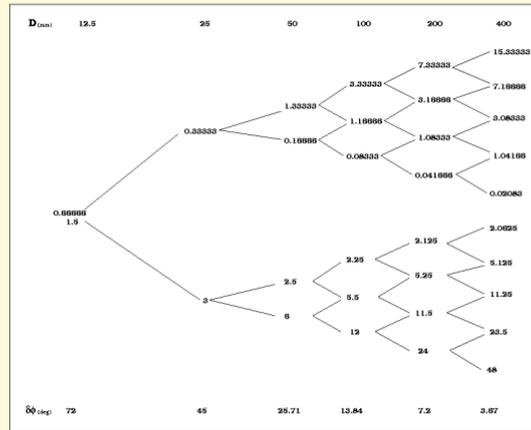
Dans le cas de 3 antennes, cela a permis de découvrir "par récurrence" une relation analytique liant certains maxima.



# Du numérique au théorique

N \ D	100	200	400	800	1600
4	b1=1.2 b2=3 $\delta\phi=4.5$	b1=3.2 b2=4 $\delta\phi=34.285$	b1=5.1+3 b2=6 $\delta\phi=2.5, 1.1$	b1=9 b2=10.125 $\delta\phi=1.8$	b1=13 b2=14.083 $\delta\phi=12.85$
5	b1=1.989 b2=2.59 b3=3.605 $\delta\phi=58.91$	b1=2.507 b2=3.318 b3=7.495 $\delta\phi=50.6$	b1=2.382 b2=3.866 b3=4.439 $\delta\phi=40.66$	b1=2.179 b2=3.533 b3=7.585 $\delta\phi=34.41$	b1=3.051 b2=5.069 b3=7.422 $\delta\phi=26.21$
6	b1=1.201 b2=1.443 b3=1.763 b4=2.146 $\delta\phi=65.28$	b1=1.398 b2=2.736 b3=3.785 b4=5.472 $\delta\phi=60$	b1=1.786 b2=5.205 b3=6.168 b4=8.193 $\delta\phi=50.8$	b1=3 b2=4.546 b3=5.4 b4=8.659 $\delta\phi=44.96$	b1=2.255 b2=3.03 b3=4.017 b4=5.384 $\delta\phi=37.79$

Résultats numériques

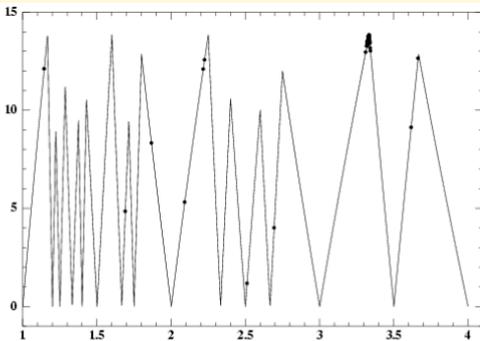


Résultats théoriques



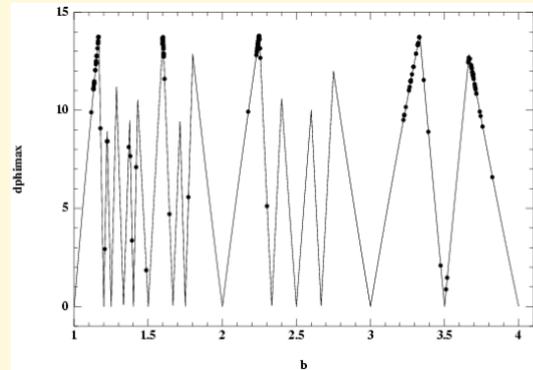
# Optimisation de radiogoniomètres

## Tous les optima par partage



Sans partage

Deux populations de 300 individus après 100 générations



Avec partage



## Partage: discussion

- Il faut fixer  $\sigma$ !
- Sélection proportionnelle (mise à l'échelle)

Instable avec tournoi

- Croisement restreint (presque) indispensable

Sélection du partenaire en fonction de la similarité

- $O(P^2)$  calculs de distance.

Mais souvent négligeable si  $\mathcal{F}$  est coûteuse

- Nécessite une grande population



## Eclaircissement

### Petrowski 96

- Fixer un rayon  $R > 0$ , et une capacité  $k$
- Trier les individus par ordre décroissant de fitness.
- Boucle sur les individus
  - Si l'individu est à distance  $> R$  des centres de niche précédents, il devient un nouveau centre de niche
  - Sinon, si la niche à laquelle il appartient a plus de  $k$  éléments, mettre sa fitness à 0.
- Sélection proportionnelle (SUS) sur les survivants.



## Eclaircissement (2)

- + simple à mettre en oeuvre
- voisinage explicite
- grande taille de population nécessaire
- +  $O(qP)$  calculs de distances ( $q$  nombre de niches)
- + robustesse (autres sélections)
- + maintient les niches plus longtemps
- + excellentes performances

**Remarque** On peut utiliser toute stratégie de sélection

Faux pour le nichage – instable avec tournoi



## Nichage : conclusion

Un gros + des algorithmes évolutionnaires

- Surcoût – calculs de distances

Pas toujours négligeable

- Dérive génétique (très) ralentie, mais inévitable

- Paramètres supplémentaires (en général)

Versions “clusterisées” des méthodes de partage

- Distances génotypiques vs distances phénotypiques

- Croisement restreint indispensable

Implicite avec les méthodes géographiques

- Indispensable pour l’optimisation multi-modale et le multi-critère



## Plan

- Techniques de nichage
- **Prise en compte des contraintes**
- Optimisation multi-objectif
- Parallélisation



## Optimisation sous contraintes

Soient  $f, (g_i)_{i=1,\dots,m} : \mathcal{S} \subset \Omega \mapsto \mathbb{R}$

$\Omega$  discret ou continu

Soit  $\mathcal{F} = \{y \in \mathcal{S} \text{ s.t. } \begin{array}{l} g_i(x) \leq 0, \quad i = 1, \dots, q \text{ et} \\ g_i(x) = 0, \quad i = q + 1, \dots, m \end{array}\}$

Find  $x^* \in \mathcal{F}$  s.t.  $f(x) = \min\{f(y); y \in \mathcal{F}\}$

- $f$  est la **fonction objectif**
- $(g_i)_{i=1,\dots,q}$ : **contraintes inégalités**
- $(g_i)_{i=q+1,\dots,m}$ : **contraintes égalité**
- $\mathcal{F}$ : **domaine faisable**
- $g_i$  est **active** au point  $x \in \mathcal{F}$  si  $g_i(x) = 0$



# Méthodes de prise en compte des contraintes

## ● Par pénalisation

Difficile à régler

- Pénalités statiques
- Pénalités dynamiques (exogènes)
- Pénalités adaptatives (endogènes)

## ● Méthodes de projection

- Contraintes linéaires : Genocop II Opérateurs spécialisés
- Opérateurs de surface Contraintes égalité, ou frontière de  $\mathcal{F}$
- Transformation de l'espace faisable Approche générale pour contraintes inégalités



## Méthodes par pénalisation

Minimiser  $\mathcal{G}(x) = f(x) + Penal(x)$

où  $Penal(X) \geq 0$ , et  
 $Penal(X) = 0$  ssi  $x \in \mathcal{F}$

- Un problème d'optimisation sans contraintes

Utilisez votre algorithme favori

- Comment choisir les termes de pénalité ?



## Types de pénalités

- **Pénalité mortelle** (Bäck et al 91):

$$Penal(X) = +\infty \text{ si } x \notin F$$

Les individus infaisables sont éliminés

Instable

- Nombre de contraintes violées

dans le cas de très nombreuses contraintes

e.g. problèmes SAT

- Mesure de violation : typiquement

$$Penal(x) = \sum_{i=1}^{i=m} \alpha_i g_i(x)^+ + \sum_{i=m+1}^{i=q} \alpha_i |g_i(x)|$$

Problème : choix de  $\alpha_i$  ?



## Pénalités statiques

$\alpha_i$  sont déterminés par l'utilisateur

- $\alpha_i$  trop petit : le minimum de  $\mathcal{G}$  peut être infaisable
- $\alpha_i$  trop grand : la recherche est restreinte à  $\mathcal{F}$

Pas de raccourcis à travers la région infaisable (cf pénalité mortelle)

### Exemple

Homaifar et al. - 94

$$Penal(x) = R(g^+(x)) [g^+(x)]^2$$

avec  $R(g^+(x)) = R_i$  si  $G_{i-1} < g^+(x) \leq G_i$   $i = 1, l$   
 $\implies m(2l + 1)$  paramètres sensibles à régler !



## Pénalités dynamiques (exogènes)

$\alpha$  est modifié suivant un schéma pré-défini

En général croissant

De très nombreuses variantes. Par exemple

Kane and Schoenauer 96

$$\alpha(t) = \alpha_0 + (\alpha_F - \alpha_0) \left( \frac{t}{T} \right)^\beta$$

où  $T$  est le nombre total de générations

### Mais

- Convergence très sensible aux paramètres. e.g.  $\alpha_F, \alpha_0, T$  et  $\beta$
- Réglage automatique d'après la population initiale ???



## Pénalités adaptatives (endogènes)

Modifier la pénalité d'après l'historique de l'évolution

- **Bean and Hadj-Alouane 92**

$$\alpha(t+1) = \begin{cases} \frac{\alpha(t)}{\beta_1} & \text{if } x_b(t), \dots, x_b(t-k) \in \mathcal{F} \\ \beta_2 \alpha(t) & \text{if } x_b(t), \dots, x_b(t-k) \notin \mathcal{F} \\ \alpha(t) & \text{otherwise} \end{cases}$$

$\beta_1, \beta_2 > 1$ ,  $x_b(t)$  meilleur à la génération  $t$ .

### **Mais**

Ne considèrent que le meilleur individu passé

Pas de garantie de faisabilité de la solution obtenue !



# Supériorité des points faisables

Deb 99

Un point faisable sera toujours préféré à un point infaisable

$$\mathcal{G}(x) = \begin{cases} f(x) & \text{if } x \in \mathcal{F} \\ r g^+(x) + \max_{x \in \mathcal{F}} \{f(x)\} & \text{otherwise} \end{cases}$$

Comparaison de 2 individus :

- Si les deux sont faisables, comparer les fonctions objectif
- Si un seul est faisable, il gagne
- Si les deux sont infaisables, comparer les violations de contraintes

Les points infaisables sont poussés vers le domaine faisable

Problème : mise à l'échelle de différentes contraintes ...



# L'algorithme adaptatif ségrégationnel: ASCHEA

S. BenHamida, 98-01

## Objectif :

Maintenir des individus faisables **et** infaisables dans la population

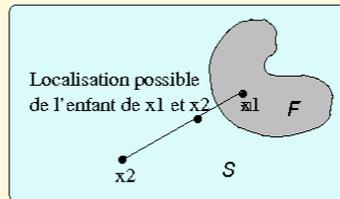
(évaluation)

### Pénalisation adaptative

si trop d'infaisables  
pénalisation ↗  
si trop de faisables  
pénalisation ↘

(croisement)

### Sélection/Séduction



(remplacement)

### Remplacement ségrégationnel

Objectif:  
augmenter la chance de survie des faisables  
(maintenir une proportion de faisabilité)



## Pénalité adaptative

**But** : maintenir une proportion  $\tau_{target}$  d'individus faisables

$$penal(\vec{x}) = \alpha(t) \left( \sum_{j=1}^q g_j^+(\vec{x}) + \sum_{j=q+1}^m |h_j|(\vec{x}) \right)$$

$\implies$  adapter  $\alpha(t)$  selon la proportion d'individus faisables  $\tau_t$

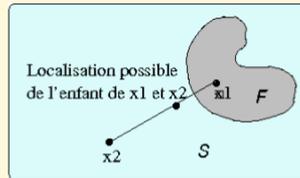
$$\begin{array}{ll} \text{si } (\tau_t > \tau_{target}) & \alpha(t+1) = \alpha(t) / \beta \\ \text{sinon} & \alpha(t+1) = \alpha(t) * \beta \end{array}$$

$\beta > 1$  et  $\tau_{target}$  : paramètres utilisateur (e.g. 1.1 et 0.6)



## Sélection/Séduction

**But :** 1- explorer la frontière de la région faisable  $\mathcal{F}$ ,  
2- attirer les individus infaisables vers  $\mathcal{F}$ .



Croisement arithmétique entre faisables et infaisables

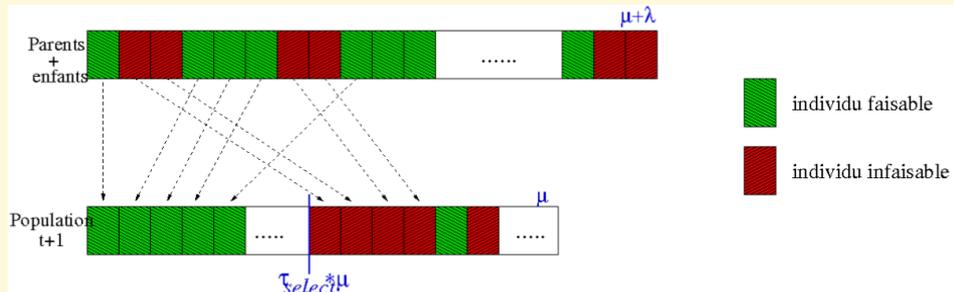
**Si**  $(0 < \tau_t < \tau_{target})$  **et**  $(x_1)$  faisable  
sélectionner  $(x_2)$  infaisable  
**sinon** sélectionner  $(x_2)$  dans toute la population



## Remplacement ségrégationnel dans un $(\mu + \lambda)$ -ES

**But** : Augmenter la chance de survie des individus faisables

1. Sélection des  $\tau_{select} * \mu$  meilleurs faisables.
2. Sélection “normale” des  $(1 - \tau_{select}) * \mu$  suivants





## ASCHEA : améliorations

- Coefficient de pénalité propre à chaque contrainte:

$$penal(\vec{x}) = \sum_{j=1}^q \alpha_j(t) g_j^+(\vec{x}) + \sum_{j=q+1}^m \alpha_j(t) |h_j|(\vec{x}),$$

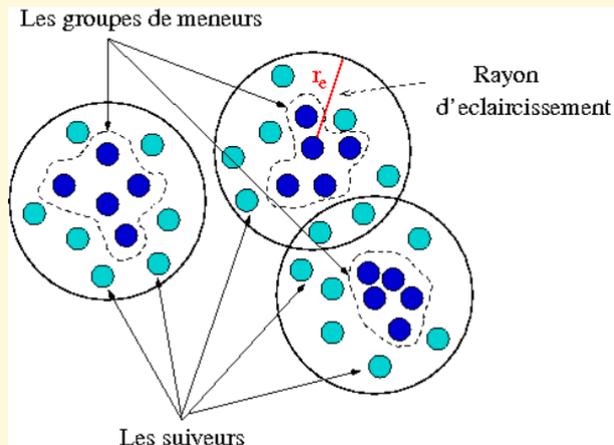
Avec adaptation commune

- Ajout de nichage (éclaircissement adapté)



## Sélection avec nichage

Éclaircissement adapté (Pétrowski, 96)

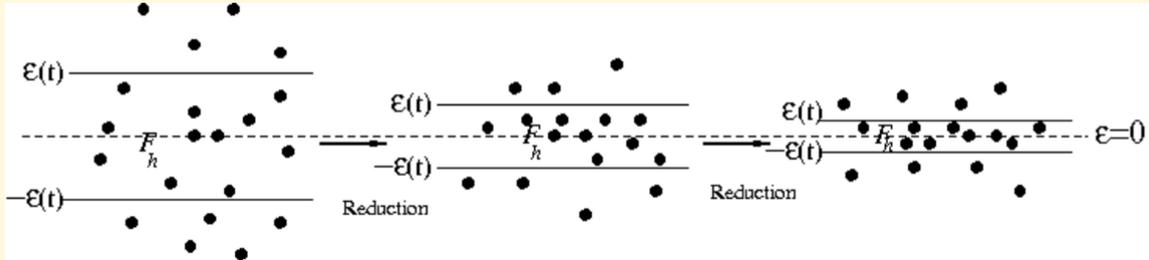


1. Classer les individus en **meneurs** et **suiveurs**
2. Appliquer la sélection ségrégationnelle sur les **meneurs**
3. Compléter si besoin avec les **suiveurs**



## Prise en compte des contraintes égalité

- Remplacer  $h_i(x) = 0$  par  $-\varepsilon_i < h_i(x) < \varepsilon_i$
- Réduire progressivement les  $\varepsilon_i$

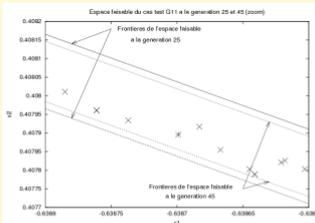


$$\begin{array}{ll} \text{si } (\tau_t(j) > \tau_{reduct}) & \epsilon_j(t+1) = \epsilon_j(t) / fact_\epsilon \\ \text{sinon} & \epsilon_j(t+1) = \epsilon_j(t) \end{array}$$

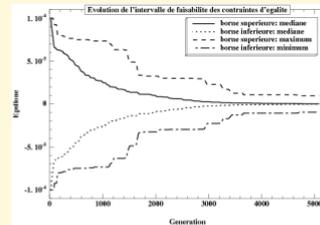


# Ajustement dynamique: Résultats

(a) Exemple de réduction de  $\mathcal{F}_h$



(b) Evolution de  $\varepsilon_t$



Résultats pour les cas test G3, G5 et G11 avec  $\tau_{reduct} = 0.6$  et  $fact_e = 1.01$

Fonction	Performance			Violation (somme des $ h_j $ )			
	Meill	Med	Moy	Min	Med	Moy	Max
G3	<b>1</b>	0.99999	0.9932	$9.18 \cdot 10^{-9}$	$1.98 \cdot 10^{-7}$	$4.37 \cdot 10^{-5}$	$2.66 \cdot 10^{-4}$
G5	5126.52	5154.13	5168.9	$1.6 \cdot 10^{-9}$	$6.18 \cdot 10^{-7}$	$2.61 \cdot 10^{-4}$	0.008
G11	<b>0.75</b>	<b>0.75</b>	0.7519	$3.19 \cdot 10^{-12}$	$2.11 \cdot 10^{-11}$	$7.43 \cdot 10^{-11}$	$9.5 \cdot 10^{-10}$



## ASCHEA : Conclusion

- fonction de pénalisation adaptative un coefficient par contrainte
- croisement entre faisables et infaisables
- remplacement ségrégationnel Simplicité de mise en oeuvre
- procédure de nichage avec rayon adaptatif
- prise en compte des **contraintes égalité** robuste si unique contrainte  
risque de perte des faisables

Résultats état de l'art sur les fonctions test usuelles

y compris contraintes égalités

Mise en oeuvre délicate

mais paramètres ( $\tau_{select}$ ,  $\beta$ ,  $\tau_{reduct}$ ) robustes



## Méthodes de projection

### Recherche dans le domaine faisable

tout en maintenant de bonnes propriétés de recherche

- Facile dans des cas particuliers      Boules, et autres espaces convexes, ...
- Post-traitement      Pénalité mortelle, réparation, ...  
Mais biaise généralement beaucoup trop la recherche
- Algorithmes spécifiques
  - Initialization      Uniforme sur  $\mathcal{F}$ ?
  - Opérateurs fermés sur  $\mathcal{F}$     i.e. les enfants de parents de  $\mathcal{F}$  sont dans  $\mathcal{F}$
  - En respectant quelques principes de base      cf Radcliffe 93
- Transformer  $\mathcal{F}$       en un espace facile à explorer



## Contraintes linéaires

... mais fitness non-linéaire !

**Genocop II** (Michalewicz & Janikow 91) :

- Contraintes égalités :  
Elimination de variables
- Contraintes inégalités :
  - Croisement arithmétique (or BLX- $\alpha$ )
  - Mutation tronquée
  - Initialisation ?

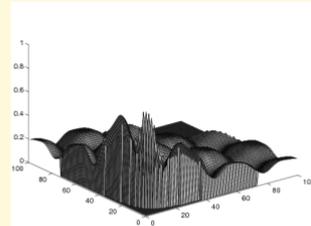
Généralisation ?



## Recherche à la frontière de l'espace faisable

un premier exemple (Michalewicz 96)

$$\begin{cases} \max f(\vec{x}) = \left| \frac{\sum_{i=1}^n \cos^4(x_i) - 2 \prod_{i=1}^n \cos^2(x_i)}{\sqrt{\sum_{i=1}^n i x_i^2}} \right|, \\ 0 \leq x_i \leq 10 \\ \prod_{i=1}^n x_i \geq 0.75 \end{cases}$$



La seconde contrainte est active à l'optimum:  
Recherche sur l'hyperboloïde  $\prod_{i=1}^n x_i = 0.75$



## Recherche sur l'hyperboloïde

- **Initialisation**

$$x_i = U[0, 10], i \in [1, n]$$
$$x_i \leftarrow \frac{0.75}{\prod_{j=1}^n x_j} x_i$$

- **Croisement "géométrique"**

$$(x_i)(y_i) \rightarrow (x_i^\alpha y_i^{1-\alpha})$$

avec  $\alpha = U[0, 1]$

- **Mutation "géométrique"**

$$x_i \rightarrow qx_i, x_j \rightarrow \frac{x_j}{q}$$

avec  $q = U[0, 1]$

Meilleurs résultats connus (jusqu'à très récemment :-)

## Recherche sur la sphère



$$\begin{cases} \text{Prod}(\vec{x}) = (\sqrt{n})^n \cdot \prod_{i=1}^n x_i, \\ 0 \leq x_i \leq 1, \\ \sum_{i=1}^n x_i^2 = 1. \end{cases}$$

- **Initialisation**

$U([0, 1]^n)$  + projection (**pas** uniforme).

- **Croisement "sphérique"**

$$(x_i)(y_i) \rightarrow (\sqrt{\alpha x_i^2 + (1 - \alpha)y_i^2})$$

avec  $\alpha = U[0, 1]$

- **Mutation "sphérique"**

$$x_i \rightarrow p \cdot x_i$$

$$x_j \rightarrow q \cdot x_j,$$

avec  $p = U[0, 1]$  et  $q = \sqrt{\left(\frac{x_i}{x_j}\right)^2(1 - p^2) + 1}$ .

Excellents résultats très rapidement



# Généralisation

## Recherche sur une surface

- Contraintes égalité,
- ou si l'on sait quelles contraintes seront actives
- Restreindre la recherche à une surface (variété Riemanienne)  $\mathcal{B}$ .
- Il faut définir des opérateurs de surface
- problème dépendant
- mais toujours possible sur la sphère



## Opérateurs de surface

- Basés sur des courbes Géométrie Riemanienne
- Intersection avec des hyper-plans Géométrie analytique
- Représentations paramétriques Operateurs sur les paramètres
- Transformer  $\mathcal{B}$  en une (partie de) sphère  
Utiliser les opérateurs "sphériques" ou tout autre comme ci-dessus  
Calcul numérique (e.g. par bisection) coûteux, facile si  $\mathcal{F}$  est convexe (ou étoilé)  
→ moins performant que des opérateurs spécifiques



## Prise en compte des contraintes : conclusions

- Difficile de comparer les diverses méthodes

Problèmes différents, différents programmes

Contraintes égalité

- Les méthodes spécifiques sont cependant toujours les meilleures

- Les méthodes par pénalisation sont un bon compromis

- Commencer par des pénalités statiques

Résoud 80% des cas !

- Pénalités adaptatives ou stochastic ranking

Vaut l'effort supplémentaire



## Plan

- Techniques de nichage
- Prise en compte des contraintes
- **Optimisation multi-objectif**
- Parallélisation



## Optimisation multi-objectif

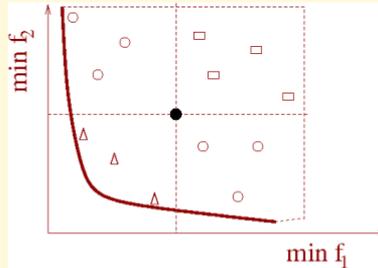
- Les problèmes réels mettent en fait en jeu plusieurs objectifs
- le plus souvent contradictoires e.g. coût et qualité
- Pas une solution, mais des compromis

Comment prendre une décision ?



## Notion de dominance

Au sens de Pareto :  $x$  domine  $y$  si  $\forall i \in [1, K], \mathcal{F}_i(x) \leq \mathcal{F}_i(y)$  et  
 $\exists i \in [1, K], \mathcal{F}_i(x) < \mathcal{F}_i(y)$



- domine les carrés, est dominé par les triangles, pas comparables avec les cercles

Front de Pareto = individus non-dominés      Meilleurs compromis possibles

Démarche idéale : Prise de décision à information complète

i.e. après identification du front de Pareto

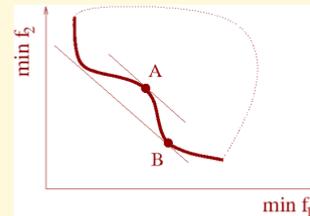
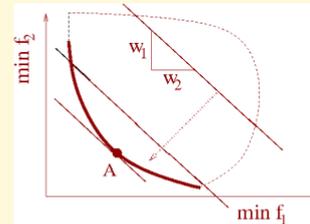


## Approches classiques

### Aggrégation des critères

$$\left\{ \begin{array}{l} \text{Argmin} \sum_0^N w_i \mathcal{F}_i(X), \\ w_i \geq 0 \text{ ssi } \mathcal{F}_i \text{ à min.} \\ \sum |w_i| = 1 \end{array} \right.$$

- Il faut choisir les poids *a priori*
- Autant de calculs que de combinaisons de poids
- Parties concaves du Front de Pareto inaccessibles

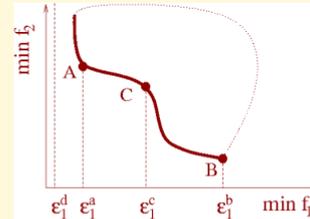




## Approches classiques

### $\varepsilon$ -contraintes

$$\begin{cases} \text{Argmin} \mathcal{F}_{i_0}(X), \\ \mathcal{F}_i(X) < \varepsilon_i \text{ pour } i \neq i_0 \end{cases}$$



- Il faut choisir les  $\varepsilon_i$  *a priori*
- Autant de calculs que de valeurs des  $\varepsilon_i$
- Mais tout le front de Pareto peut être atteint



## Approches évolutives non-Pareto

- **VEGA**

- Sélectionner  $\frac{P}{k}$  individus selon chacun des  $k$  critères
- Ne conserve que les points extrêmes pour un des critères.

- **Poids mouvants**

- Poids (normalisés) codés dans le génotype
- Partage et croisement restreints indispensables

- Peu efficaces

Performances améliorées par archivage



## Approches évolutionnaires Pareto (EMOAs)

Il “suffit” de modifier la sélection . . .

Se ramener à un scalaire?

et on peut espérer échantillonner tout le front de Pareto

génotypiquement ou phénotypiquement?

Mais attention à la perte de diversité

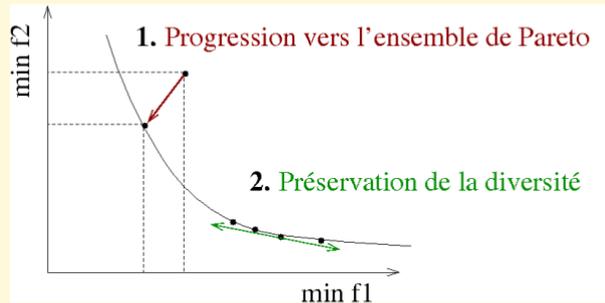
Dérive génétique

### Historique

- MOGA Fonseca et Fleming 1993
- NSGA, puis NSGA-II Deb et al. 1993 puis 1999
- NPGA Horn et Nafpliotis 1993
- SPEA, puis SPEA2 Eckart et Zitzler  
1998, avec introduction d’une archive, puis 2000
- PAES, puis PESA Knowles et Corne 1999, puis 2001



## Sélection multi-objectif



### Comparaison de deux individus

1. **Dominance** au sens de Pareto  
Ordre seulement partiel : si non-comparables
2. **Densité** de solutions

**Elitisme indispensable** Généralement via une archive

Echantillonnage d'une approximation du front de Pareto

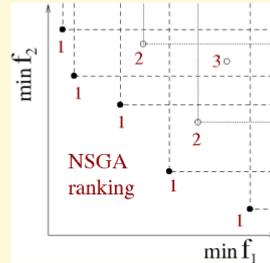


## Un exemple : NSGA-II

### Rang de Pareto

Non-dominés = rang 1

Non-dominés sauf par rang 1 = rang 2 ...



### Mesure de surpeuplement (Crowding distance)

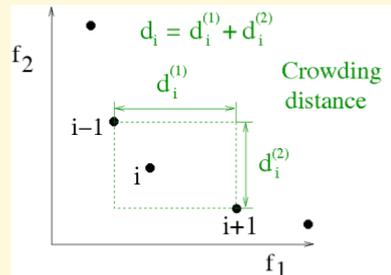
Pour chaque critère  $c$

- Trier selon critère  $c$

cher?

- $d_c(x_i) = d(x_i, x_{i-1}) + d(x_i, x_{i+1})$ ,

$$d_{\text{surpeuplement}} = \sum_c d_c(x)$$



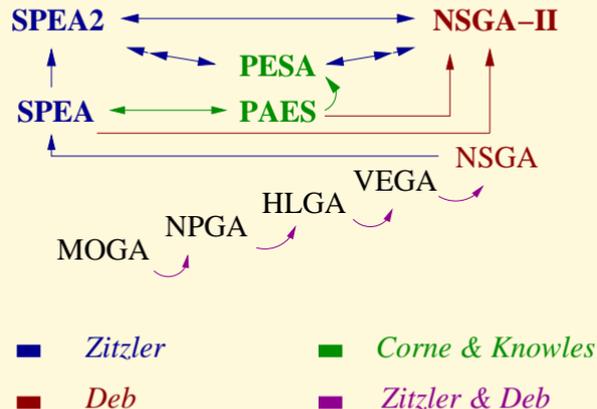
**Elitisme:**  $(\mu + \mu) - ES$



## Comparaisons

### Critères de comparaisons ?

- Précision de l'approximation du front quand il est connu
- Comparaison des fronts obtenus en un temps donné au moins 16 mesures répertoriées



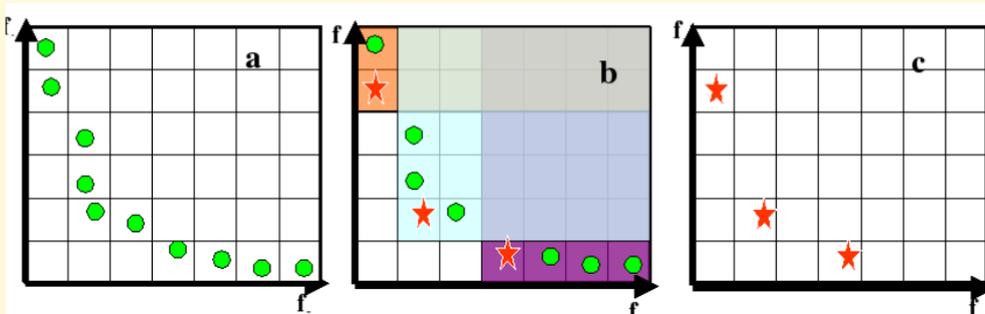


# $\varepsilon$ -MOEA

Deb et al., 2003

$\varepsilon$ -dominance: dominance à  $\varepsilon$  près

Laumans, 2002



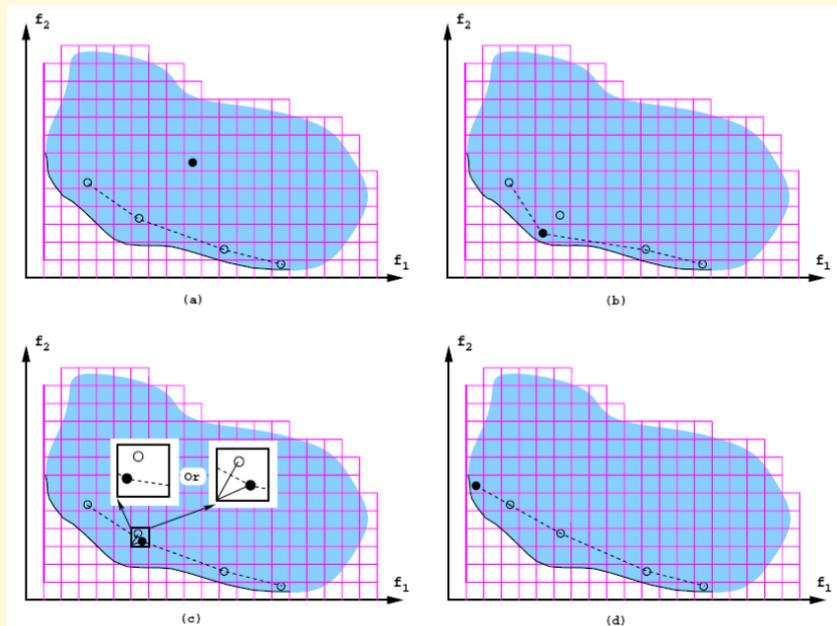
Les trois étoiles  $\varepsilon$ -dominent les autres points



## $\varepsilon$ -MOEA (2)

Algorithme de type **steady-state**,  
maintient une population  $\Pi$  et une archive  $\mathcal{A}$

1. Initialisation de  $\Pi$ , copie des non-dominés dans  $\mathcal{A}$
2. Choisir un parent dans  $\Pi$  2-tournoi, dominance, uniforme sinon
3. Choisir l'autre parent dans  $\mathcal{A}$  uniforme
4. Créer un enfant Croisement, mutation
5. L'insérer dans  $\Pi$  basé sur la dominance, uniforme sinon
6. L'insérer dans  $\mathcal{A}$   $\varepsilon$ -dominance
7. Goto 2



- (a) : rejeté ; (b) : remplace l' $\epsilon$ -dominé  
 (c) remplace, car domine, ou est plus proche de l'origine ;  
 (d) nouveau point  $\epsilon$ -extrémal

Jamais plus d'un individu par hyper-rectangle



## $\epsilon$ -MOEA Performances

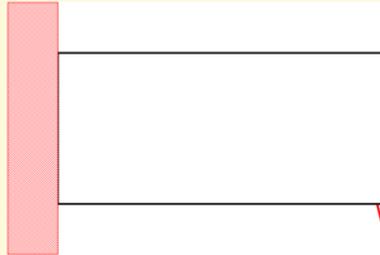
- Temps calcul très réduit
    - 10 fois moins que NSGA-II
    - 20 fois moins que PESA
    - 100+ fois moins que SPEA
  - Approximation du front de Pareto très bonne  
bonne quelquefois
  - Recouvrement excellent
- Clustering
- Très légèrement moins
- moins bon que SPEA2

Un excellent compromis



## Exemple : Optimisation de formes multi-objectif

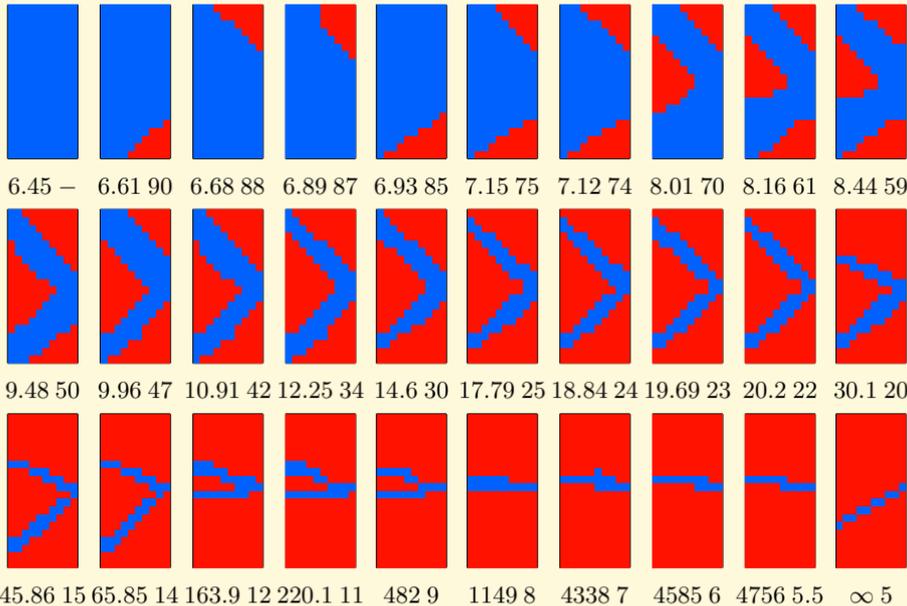
H. Hamda et O. Roudenko 2000



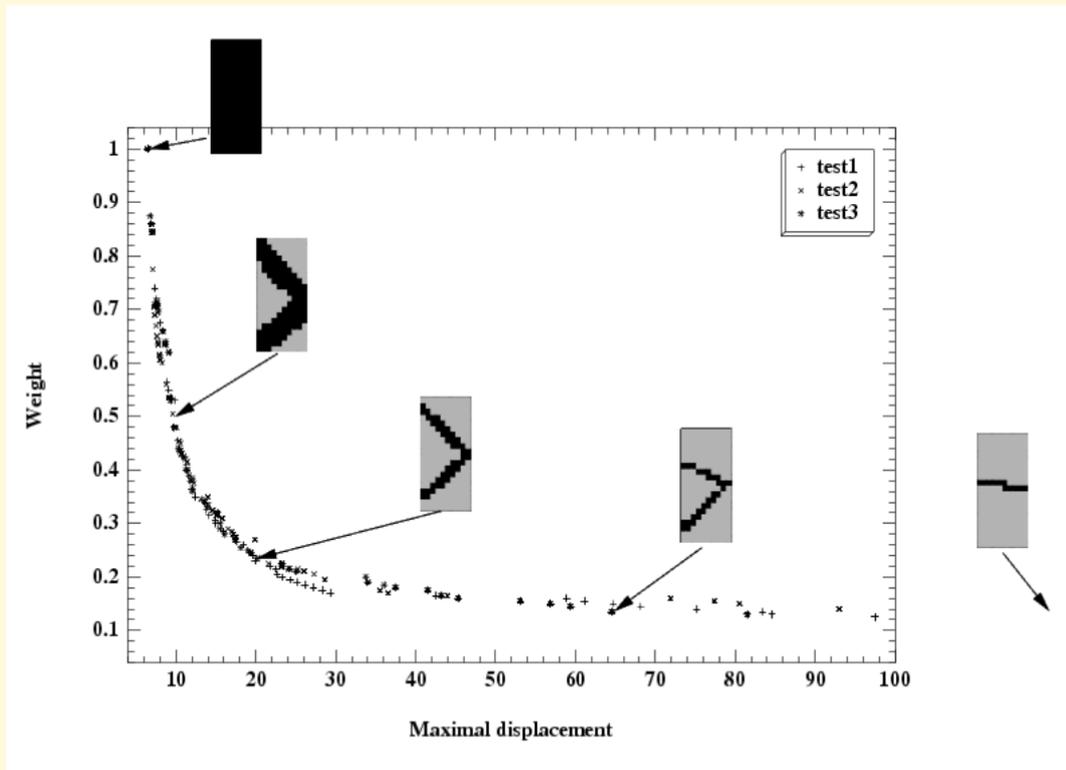
- Trouver une forme dans un domaine donné
- minimisant le poids
- et le déplacement maximal sous une charge donnée  
ou maximisant la première valeur propre, ou ...
- Représentation par diagrammes de Voronoi
- NSGA-II, taillePop = 300



# Poids et déplacement maximal – Console 10x20

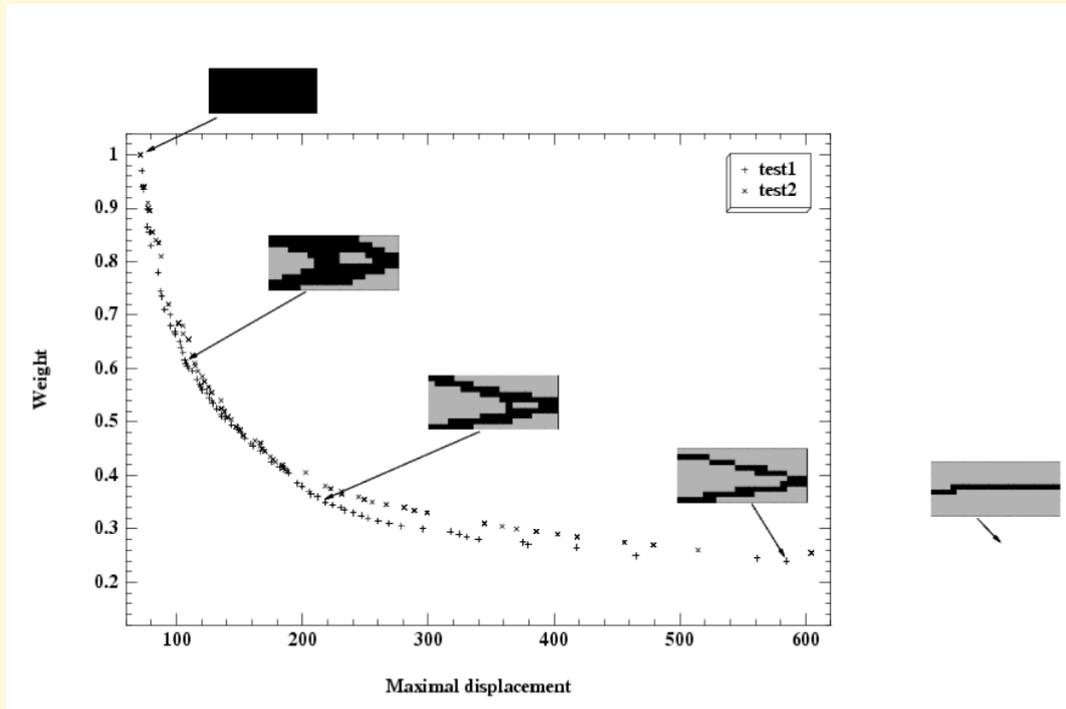


**CPU time  $\equiv$  single objective!**



Trois fronts de Pareto indépendants (400 générations de 300 individus)

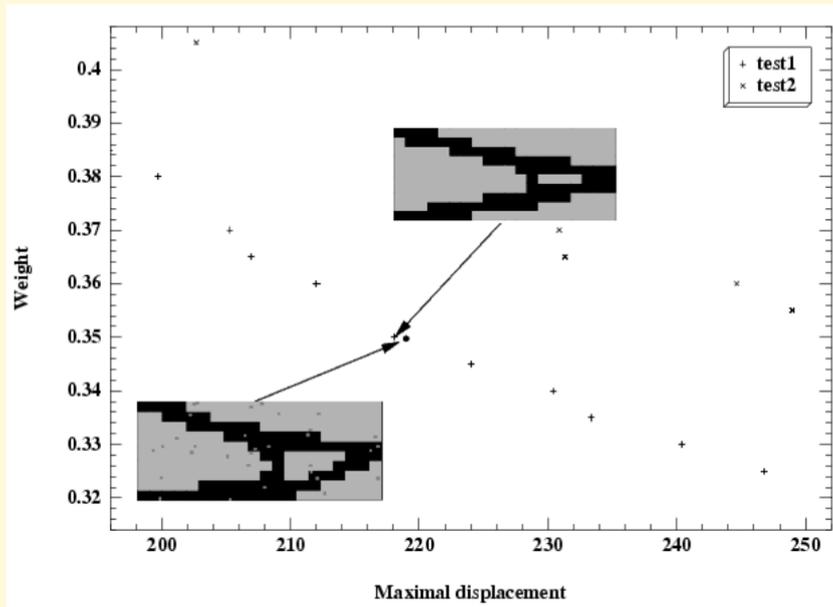
# Poids et déplacement maximal – Console 20x10



Deux fronts de Pareto indépendants (400 générations de 300 individus)



# Multi-objectif vs $\varepsilon$ -contraintes (Console 20x10)



Front de Pareto au voisinage du déplacement maximal 220.

Haut ← multi-objectif

Bas ← contrainte



## Multi-objectif : conclusions

- Un champ de recherche très actif Conférence EMO bisannuelle
- où les AEs n'ont pas beaucoup de concurrents tout le front de Pareto
- Meilleur choix ?
  - NSGA2, le plus robuste base de comparaison
  - MO-CMA-ES, le plus précis variables continues
  - $\varepsilon$ -MOEA pour des fonctions objectif chères
- **mais**
  - attention aux critères de comparaison Hypervolume
  - mauvaises performances pour plus de 3 objectifs Teytaud, 2007
    - **Many Objective Optimization**



## Plan

- Techniques de nichage
- Prise en compte des contraintes
- Optimisation multi-objectif
- **Parallélisation**



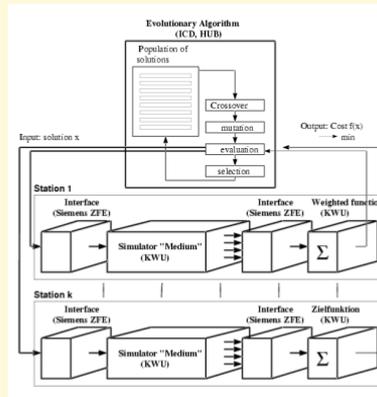
## Parallélisation

- Modèle maître-esclave
  - Synchrone ou asynchrone
  - Etude théorique pour le modèle synchrone
- Modèle en îlots
  - Migration entre îlots
  - Topologie ?
- Modèle totalement distribué
  - Sur machines massivement parallèles
  - voire sur machines séquentielles ...



# Approche maître/esclave

Seul le calcul de la performance est distribué



## Synchrone

- Génération par génération
- Envoi de PopSize/NPROC calculs à chaque processeur
- Dynamique **inchangée**

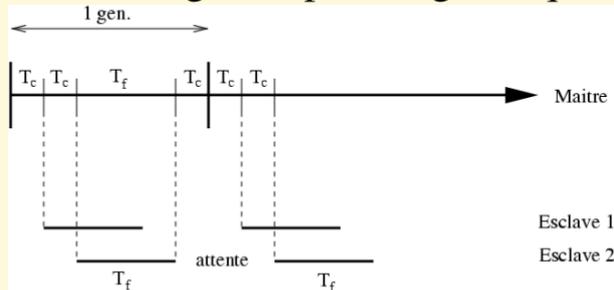


## Approche maître/esclave : Analyse d'efficacité

E. Cantu-Paz

Efficient and accurate parallel GA, Kluwer, 2000

**Hypothèses** : Calculs homogènes, partie “génétique” négligeable



- Nombre de processeurs  $\mathcal{P} = \mathcal{S} + 1$
- Temps de calcul par individu  $T_f$
- Temps de calcul par processeur  $\frac{nT_f}{\mathcal{P}}$

$$\text{Temps total : } T_P = \mathcal{P}T_c + \frac{nT_f}{\mathcal{P}}$$



## Analyse d'efficacité (2)

**Nombre optimal de processeurs** :  $\mathcal{P}^* = \sqrt{\frac{nT_f}{T_c}}$

**Raffinement** :  $T_c = Bx + L$

- $B$  dépend de la bande passante du réseau
- $x$  est la quantité d'information
- $L$  (*latence*) dépend du système

Alors,  $T_{envoi} = B \frac{nl_i}{\mathcal{P}} + L$ ,  $T_{rec.} = B \frac{nl_f}{\mathcal{P}} + L$ , et

$$T_P = (\mathcal{P} - 1)T_{envoi} + T_{rec.} + \frac{nT_f}{\mathcal{P}} \quad \text{et donc}$$

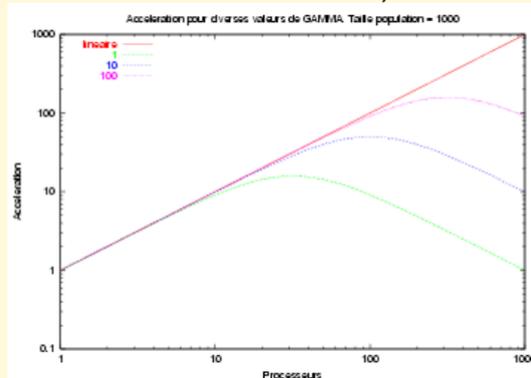
$$\mathcal{P}^* = \sqrt{\frac{n(T_f + B(l_f - l_i))}{L}}$$

Équivalent lorsque  $B \rightarrow 0$



## Accélération

Comparaison avec le séquentiel :  $\frac{T_S}{T_P} = \frac{nT_f}{\frac{nT_f}{\mathcal{P}} + \mathcal{P}T_c} = \frac{n\gamma}{\frac{n\gamma}{\mathcal{P}} + \mathcal{P}} \stackrel{?}{>} 1$



Accélération ssi

$$\gamma = \frac{T_f}{T_c} > \frac{\mathcal{P}^2}{n(\mathcal{P} - 1)}$$

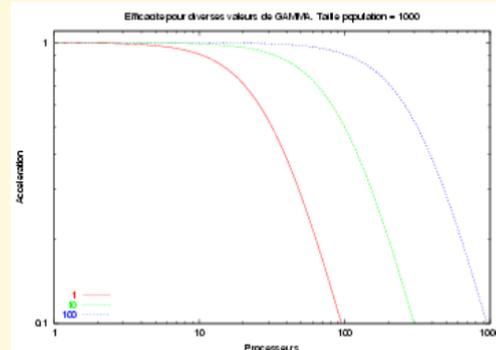
Par exemple, pour  $\mathcal{P} = 2$ , il faut  $\gamma > \frac{4}{n}$



# Approche maître/esclave

## Efficacité

$$E_P = \frac{\text{Accel.}}{\mathcal{P}} = \frac{n\gamma}{n\gamma + \mathcal{P}^2}$$



Pour une efficacité donnée

$$\mathcal{P}_E = \sqrt{\frac{1-E}{E}} \sqrt{\frac{nT_f}{T_c}} = \sqrt{\frac{1-E}{E}} \mathcal{P}^*$$

Conséquence :  $\mathcal{P}_c = \mathcal{P}^* \Leftrightarrow E = \frac{1}{2}$

**On doit pouvoir faire mieux !**



## Approche maître/esclave

### Conclusions modèle synchrone

**Experiences** E. Cantu-Paz 1999 –  $T_c = 19ms$ ,  $n = 120$   
Accord parfait avec la théorie

### Modèle M/E synchrone: Conclusions

- Accélération quasi-linéaire (presque) garantie

Pour des calculs de performance longs

Et pas trop de processeurs

- Efficacité réduite

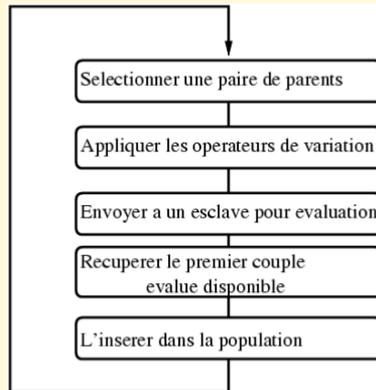


# Approche maître/esclave

## Modèle asynchrone

### SSGA asynchrone

- Un (ou deux) individu à chaque fois
- Évalué par un processeur esclave
- Premier revenu, premier “servi”





## Approche asynchrone (2)

### Remplacement

- Aléatoire
- Basé sur la performance Le plus mauvais, ou par tournoi
- Basé sur l'âge Le plus vieux, ou par tournoi

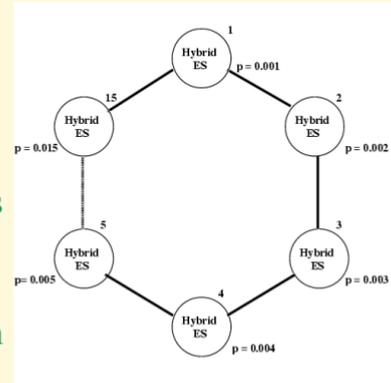
### Accélération et efficacité

- Dynamique à peine différente
- Tolérance aux pannes
- Robuste en environnement hétérogène ou performances non-linéaires
- Supporte plus de processeurs Attention au trafic au noeud maître
- Meilleure efficacité Possibilité de "bufferiser" les individus
- ... expérimentale Étude théorique plus délicate



## Modèle en îlots

- Population indépendantes (*demes*)  
Eventuellement, différents paramètres
- Algorithms standard + **Migration**  
Ou polénisation



- Paramètres de migration Topologie, fréquence ?
  - Analyse d'efficacité plus délicate Comparaison à précision identique
- E. Cantu-Paz, 2000 sur les chaînes de bits



## Modèle en îlots (2)

### Paramètres de migration

- Choix des migrants Au hasard, ou les meilleurs
- Choix des remplacés Au hasard, les plus mauvais, ou les plus proches

### Influence sur la pression de sélection

**Critère** : Temps mis par le meilleur pour envahir la population

#### Hypothèses

- Algorithme limité à la sélection – 2 types d'individus seulement
- $P_t$  proportion de bons au temps  $t$  ( $P_0 = \frac{1}{n}$ )
- Tournoi de taille  $s$  :  $P_{t+1} = 1 - (1 - P_t)^s$
- Fraction  $\rho$  de la population migre



## Modèle en îlots

### Pression de sélection

- Choix et remplacement random

Rien ne change

- Choix des bons, remplacement des mauvais

$$P_{t+1} = 1 - (1 - P_t)^s + \rho$$

- Choix des bons, remplacement random

$$P_{t+1} = 1 - (1 - P_t)^s + \rho(1 - P_t)^s$$

- Choix random, remplacement des mauvais

$$P_{t+1} = 1 - (1 - P_t)^s + \rho(1 - (1 - P_t)^s)$$



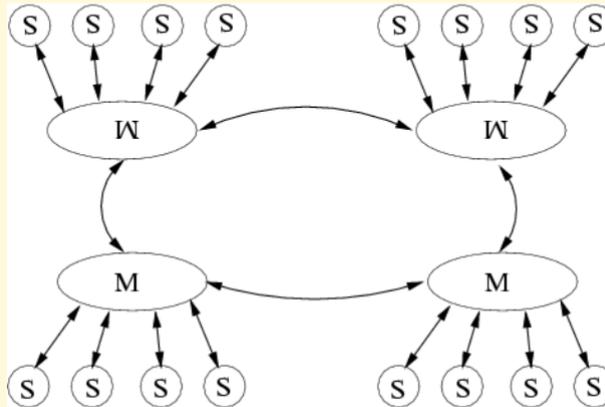
## Modèle en îlots

### Topologie des migrations

- Totalemment connectés
- Degré du graphe (nombre de voisins) Pas un critère discriminant
- Résultats théoriques ... sur une génération  
E. Cantu-Paz, 2000 sur les chaînes de bits



“Mélange” des deux modèles précédents:



- Calcul des performances “localement distribuées”
- Migrations entre les îlots
- Pas d’analyse d’efficacité

Synchronisation?



## Populations totalement distribuées

- Un individus par noeud
- Sur machines massivement parallèles
- Sélection locale
- Croisement local
- Des niches spatiales apparaissent
- Dynamique particulière Accélérations quelquefois supra-linéaires
- Quelques résultats théoriques récents Robustesse accrue face au bruit G. Rudolph, 1999
- ... utilisé quelquefois sur machines séquentielles !



## Parallélisation : conclusions

- Modèle maître-esclave : rien à perdre
- Modèle en îlot : mise en oeuvre plus délicate  
Bien adapté aux P2P
- Population distribuée : sur machines séquentielles

Mise en oeuvre facile