

TECHNISCHE UNIVERSITÄT DORTMUND

Wintersemester 2007/08

Praktische Optimierung
(Vorlesung)

Prof. Dr. Günter Rudolph
Fakultät für Informatik
Lehrstuhl für Algorithm Engineering (LS XI)
Fachgebiet *Computational Intelligence*

MOP: Metaheuristiken – Multipunktmethoden

Bisher:
je Iteration wird ein Entscheidungsvektor variiert und bewertet

Jetzt:
je Iteration werden mehrere Entscheidungsvektoren variiert und bewertet

- **Historisches:**
VEGA, VOES, MOGA, PPES
- **Aktuelles:**
NSGA-2, SPEA-2, SMS-EMOA, MSOPS

} unvollständige Liste

} z.Z. am häufigsten eingesetzt

Rudolph: PO (WS 2007/08) • MOP: Metaheuristiken – Multipunktmethoden 2

MOP: Metaheuristiken – Multipunktmethoden

Schaffer 1984: **VEGA** (vector-evaluated genetic algorithm)

d Ziele → d Teilpopulationen P_1, \dots, P_d der Größe s

Initialisiere alle Individuen zufällig

repeat

$\forall i : \forall x \in P_i : \text{fitness}(x) = f_i(x)$

je Teilpopulation P_i :

- wähle s Individuen (auch mehrfach) durch proportionale Selektion
- wende Crossover und Mutation an

ergibt $d \cdot s$ Nachkommen

alle Nachkommen zufällig in d Teilpopulationen aufteilen

until Stoppkriterium erfüllt

Rudolph: PO (WS 2007/08) • MOP: Metaheuristiken – Multipunktmethoden 3

MOP: Metaheuristiken – Multipunktmethoden

Kursawe 1990: **VOES** (vector optimization evolution strategy)

d Ziele → diploide bzw. polyploide Chromosomen

Idee: „Reservelager“ für Umweltveränderungen

Ersatzzielfunktion mit Wichtungsvektor $w = (w_1, \dots, w_d)$ fix oder zufällig ändernd

VOES (skizziert)
Mutation + Rekombination „wie üblich“
Austausch dominanter Gene mit rezessiven Genen mit W'keit 1/3
($\mu+\lambda$)-Selektion via Ersatzzielfunktion
neue Eltern ins Archiv
Archiv filtern (nur nicht-dominierte Elemente)

Rudolph: PO (WS 2007/08) • MOP: Metaheuristiken – Multipunktmethoden 4

Fonseca & Fleming 1993: **MOGA** (multi-objective genetic algorithm)

→ Fitness abh. von Rang und Einnischungsmaß

- Rangzuordnung:
 $\text{Rang}(x) = 1 + \text{card}(\{y \in \text{Pop}: f(y) \preceq f(x)\})$
 → wenn $f(x) \in A_1$ dann $\text{Rang}(x) = 1$
- Einnischungsmaß (sharing distance)

$$N(x) = \sum_{y \in A_i} \left[1 - \left(\frac{d(x,y)}{\sigma} \right)^\alpha \right] \cdot \mathbf{1}_{[d(x,y) \leq \sigma]}$$

$d(x,y)$ = Abstand im Zielraum
 σ = Einnischungsradius
 α = Parameter > 0

Fonseca & Fleming 1993: **MOGA** (multi-objective genetic algorithm)

$$\text{fitness}(x) = \frac{F(x) N(x)}{\sum_y F(y)/N(y)}$$

Hilfsmittel: Erzeugen einer Antikettenhierarchie

→ Grundlage vieler Verfahren!

Satz: (Dilworth 1950)

Sei (F, \preceq) eine partiell geordnete Menge der Höhe h . Dann existiert eine Partition (F_1, F_2, \dots, F_h) von F aus Antiketten F_i .

Beweis: (konstruktiv)

Sei $F_1 = M(F, \preceq)$, also die Menge minimaler Elemente von F .

Setze $F_i = M(F \setminus (F_1 \cup F_2 \cup \dots \cup F_{i-1}), \preceq)$ für $i = 2, \dots, h$. ■

Goldberg 1989: **NDS** (non-dominated sorting)

Grundidee: Population aufteilen in Antikettenhierarchie F_1, \dots, F_h

erzeuge Anfangspopulation per Zufall
repeat
 bestimme d Zielwerte je Individuum
 Einteilung der Individuen gemäß Antikettenhierarchie
 Zuordnung: Individuum x hat Rang i , falls $f(x) \in F_i$
 rangbasierte Selektion
 Crossover & Mutation
until Stoppkriterium erfüllt

h: Höhe der partiell geordneten Menge

→ anscheinend nie implementiert → aufgegriffen von NSGA

(Srinivas & Deb 1994)

Laumanns et al. 1998: **PPES** (predator/prey evolution strategy)

Idee: Räuber/Beute-Ansatz

- Nachbarschaftsstruktur:
Individuen/Beutetiere leben auf Ecken eines Graphen;
Nachbarschaft definiert durch Kanten des Graphen
→ üblich: Torus
- je Ziel $r \geq 1$ Räuber
Räuber geht Zufallsweg über Graphen
frisst Beute mit schlechtester Fitness bzgl. Räubers Ziel
- Lücke auffüllen durch Rekombination/Mutation in Nachbarschaft

Laumanns et al. 1998: **PPES** (predator/prey evolution strategy)

Warum Torus?

- Räuber sollte jede Beuteposition gleichhäufig sehen
- z.B. m -reguläre nicht-bipartite Graphen
- max. erwartete Zeit bis Räuber jede Beute 1x besucht hat:
entspricht der sogenannten *cover time* $C(n)$;
bei eckensymmetrischen Graphen: $(n-1) \log n \leq C(n) \leq 2(n-1)^2$

Variationen:

- Räuber hat andere Graphstruktur als Beute
- Räuber nicht zufällig → in Richtung der schlechtesten Fitness
- Räuber hat gewichtete Ziele (Deb 2001)

bisher: „historische“ Verfahren

jetzt: aktuelle Verfahren

NSGA-2
SPEA-2
SMS-EMOA
MSOPS

Deb et al. 2000: **NSGA-2** (nondominated sorting GA, version 2)

```
erzeuge Eltern  $P(0)$  und Nachkommen  $Q(0)$ ;  $t = 0$ 
repeat
  erzeuge Antikettenhierarchie von  $R(t) = P(t) \cup Q(t)$ 
   $P(t+1) = \emptyset$ ;  $i = 1$ 
  while  $\text{card}(P(t+1) \cup R_i) \leq N$  do
     $P(t+1) = P(t+1) \cup R_i$ ;  $i++$ 
  od
  ggf. 'crowding-sort' auf  $R_i$ ; einfügen in  $P(t+1)$ 
  erzeuge Nachkommen  $Q(t+1)$  aus  $P(t+1)$ 
```

Deb et al. 2000: **NSGA-2** (nondominated sorting GA, version 2)

crowding-sort:

Sortieren der Individuen einer Front F_i bzgl. 'crowding distance'

crowding distance: („Besiedelungsdistanz“)

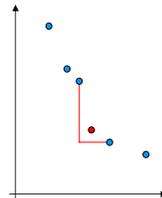
Sei $X_i = \{x \in P(t): f(x) \in F_i\}$.

$\forall k = 1, \dots, d:$

Sortiere alle $x \in X_i$ bzgl. $f_k(x)$: ergibt $x_{(1)}, \dots, x_{(\mu)}$.

Setze $d_c(x_{(1)}) = d_c(x_{(\mu)}) = \infty$.

$\forall i = 2, \dots, \mu - 1: d_c(x_{(i)}) = \frac{1}{2} (f_k(x_{(i+1)}) - f_k(x_{(i-1)}))$



Deb et al. 2000: **NSGA-2** (nondominated sorting GA, version 2)

Auswahl der Eltern zur Rekombination:

je Elter einmal 'crowded tournament selection' durchführen

crowded tournament selection:

1. falls $\text{Rang}(x) < \text{Rang}(y)$ dann selektiere x
2. falls $\text{Rang}(x) > \text{Rang}(y)$ dann selektiere y
3. falls $d_c(x) > d_c(y)$ dann selektiere x sonst y

$d_c(x)$ = crowding distance (etwa: „Besiedelungsabstand“)

Zitzler et al. (2002): **SPEA-2** (Strength Pareto EA, Version 2)

- Population $P(t)$ und Archiv $A(t)$ mit $|A(t)| = N > 1$

- „Stärke“ von Individuum x :

$$\text{strength}(x) = |\{y \in P(t): f(x) \prec f(y)\}| + |\{y \in A(t): f(x) \prec f(y)\}|$$

= Anzahl dominierter Individuen in Population und Archiv

- „Grundfitness“ (raw fitness) von Individuum x :

$$\text{raw}(x) = \sum_{\substack{y \in P(t) \\ f(y) \prec f(x)}} \text{strength}(y) + \sum_{\substack{y \in A(t) \\ f(y) \prec f(x)}} \text{strength}(y)$$

⇒ nur nicht-dominierte Individuen x haben $\text{raw}(x) = 0$

Zitzler et al. 2002: **SPEA-2** (Strength Pareto EA, Version 2)

Problem:

Viele Individuen mit gleicher Grundfitness (z.B. wenn alle nicht-dominiert)

⇒ additives Maß $\text{density}(x) \in (0,1)$ zur Differenzierung

Sei $y \in P(t) \cup A(t) \setminus \{x\}$ das Individuum mit dem k -größten euklidischen Abstand zu x .

$$\text{Dann ist } \sigma^{(k)}(x) = \|x - y\|_2 \text{ und } 0 < \text{density}(x) = \frac{1}{\sigma^{(k)}(x) + 2} \leq \frac{1}{2} < 1.$$

Schließlich:

$$\text{fitness}(x) = \text{raw}(x) + \text{density}(x)$$

Nur nicht-dominierte Individuen und nur diese haben $\text{fitness} < 1$

Zitzler et al. 2002: **SPEA-2** (Strength Pareto EA, Version 2)

$k = \lceil (|P(t)| + |A(t)|)^{1/2} \rceil$ für k-ter Nachbar

$\forall x \in P(t)$ und $x \in A(t)$: berechne **fitness**(x)

Alle Individuen mit fitness < 1 ins neue Archiv A(t+1)

Falls Archiv noch nicht voll

dann auffüllen mit Individuen mit kleinsten **fitness**-Werten ≥ 1

sonst falls Archiv überfüllt

dann iterativ ausdünnen mit der Regel:

entferne Individuum y falls $y \leq_s x$ für alle $x \in A(t+1)$

$y \leq_s x =_{\text{def}} \forall k: \sigma^{(k)}(y) = \sigma^{(k)}(x)$ oder $\exists j: \sigma^{(j)}(y) < \sigma^{(j)}(x): \forall k \neq j: \sigma^{(k)}(y) = \sigma^{(k)}(x)$

Emmerich, Beume, Naujoks 2005: **SMS-EMOA**

Fakten:

- EA muss etwas sinnvolles mit unvergleichbaren Lösungen machen.
- Gewünscht sind Lösungen nahe der Paretofront.
- Gewünscht ist eine gute Repräsentation der Paretofront.

Idee:

Nutze S-Metrik als Selektionskriterium!

⇒ erfüllt alle Forderungen!

Emmerich, Beume, Naujoks 2005: **SMS-EMOA**

Initialisiere Population von μ Individuen

repeat

wähle zwei Individuen zufällig gleichverteilt

rekombinieren und mutieren

ersetze das Individuum mit geringstem S-Metrik-Beitrag

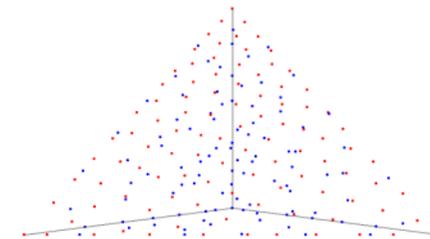
until 1 Stoppkriterium erfüllt

Berechnungskomplexität:

S-Metrik muss μ mal für μ Individuen berechnet werden

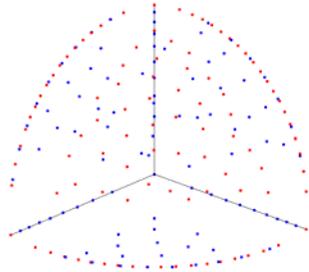
⇒ $O(\mu^{d/2+1} \log \mu)$ via Overmars/Yap (Beume, März 2006)

Emmerich, Beume, Naujoks 2005: **SMS-EMOA**



Resultate für Problem DTLZ1 ($d = 3, n = 20$)

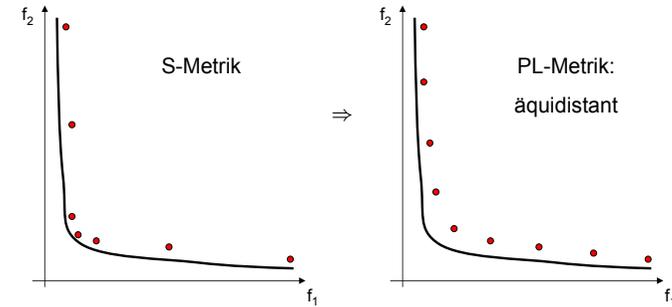
Emmerich, Beume, Naujoks 2005: SMS-EMOA



Resultate für Problem DTLZ2 (d = 3, n = 20)

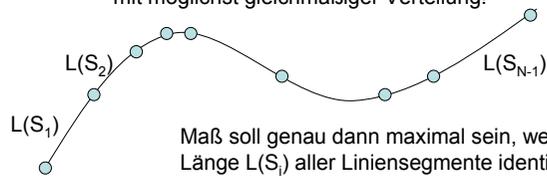
Diskussion:

Liefert max. S-Metrik wirklich geeignete Repräsentation der Paretofront?



d = 2

Gesucht: N Punkte auf der Kurve mit möglichst gleichmäßiger Verteilung!



Maß soll genau dann maximal sein, wenn Länge $L(S_i)$ aller Liniensegmente identisch!

⇒ bekanntes Optimierungsproblem!

$$\prod_{i=1}^{N-1} (1 + L(S_i)) \rightarrow \max$$

$$\sum_{i=1}^{N-1} L(S_i) = L$$

Lösung: $\forall i : L(S_i) = \frac{L}{N-1}$

$$\prod_{i=1}^{N-1} (1 + L(S_i)) \leq \left(1 + \frac{L}{N-1}\right)^{N-1} \leq \exp(L) \quad \left| \log \right.$$

$$\sum_{i=1}^{N-1} \log(1 + L(S_i)) \leq L \quad \left| : L \right.$$

PL-Metrik: $0 \leq m_{PL} = \frac{1}{L} \sum_{i=1}^{N-1} \log(1 + L(S_i)) \leq 1$

PL = Path Length

T. Wagner, J. Mehnert, GR (März 2006)

PL-Metrik:
$$D \leq mp_{PL} = \frac{1}{L} \sum_{i=1}^{N-1} \log(1 + L(S_i)) \leq 1$$

Probleme:

- 👉 Benötigt Länge L der Kurve!
 - 👉 $\forall i$: Berechne $L(S_i)$ aller Kurvensegmente
- } Paretofront muss
- bekannt sein und
 - eine Länge haben!

- 👉 Paretofront nur approximiert

⇒ **Lösung:**

definiere ein $\varepsilon > 0$;
falls Abstand zwischen i -ten Punkt und Paretofront $> \varepsilon$ dann $L(S_i) = 0$

- 👉 funktioniert nur für zwei Ziele

Hughes 2003/2005: **MSOPS** (multiple single objective Pareto sampling)

N Gewichtsvektoren $w \in W$ mit $\dim(w) = d$

Empfehlung: $1,5 N \leq \mu \leq 2 N$, wobei μ Populationsgröße

$P(t) = P(t-1) \cup \lambda$ Nachkommen

$\forall x \in P(t): \forall w \in W: \text{score } s_x(w) = \max \{ w_i f_i(x) : i = 1, \dots, d \}$

→ ergibt 'score matrix' S

ersetze Werte je Spalte durch Ränge

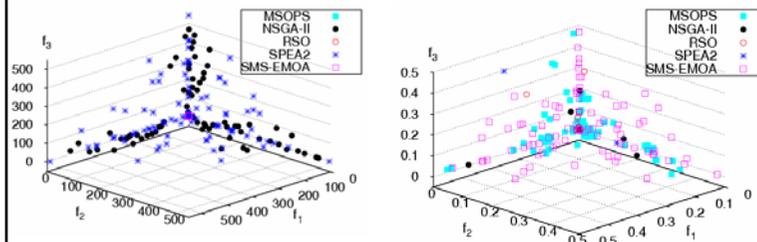
→ ergibt Matrix Q

(in jeder Zeile stehen Ränge eines x für verschiedene Gewichte w)

wähle μ beste Individuen gemäß lexikographischer Ordnung von Q

Effektivität der Verfahren für $d > 4$? (Ergebnis nach gleichem Aufwand)

Vergleich von Resultaten für Problem DTLZ1 ($d = 6, n = 30$)



Anmerkung: typische Werte; nur 3 Dimensionen aufgetragen

Offene Fragen (2006):

NSGA-2 und SPEA-2 nur effektiv bis $d \leq 3$?

MSOPS und SMS-EMOA effektiv für $d > 3$?

} Ja! Bestätigt durch ≥ 3 num. Studien (Stand 2008)