

**Własności eksploatacyjne i eksploracyjne
algorytmów ewolucyjnych
z mutacją α -stabilną**

Andrzej Obuchowicz i Przemysław Prętki

Uniwersytet Zielonogórski

Plan

- ✓ motywacja;
- ✓ rozkłady α -stabilne;
- ✓ wielowymiarowe mutacje α -stabilne;
 - ✓ nieizotropowość;
 - ✓ efekt martwego otoczenia;
- ✓ izotropowe mutacje α -stabilne;
 - ✓ wskaźniki progresu dla zmodyfikowanych strategii ewolucyjnych;
 - ✓ eksploracja kontra eksploatacja;
- ✓ podsumowanie i aktualne kierunki badań

Mutacja

Mutacja w algorytmach ewolucyjnych (i innych algorytmach stochastycznych) parametrycznej optymalizacji globalnej

$$x'_i = x_i + N(0, \sigma)$$

Uzasadnienie

- ×** o mutacji w naturze decydują czynniki fizyczne i chemiczne, które są niezależne i identycznego rozkładu;
- ×** Centralne Twierdzenie Graniczne – konieczność spełnienia warunku Lindeberga;
- ×** w przeciwnym razie – Uogólnione Centralne Twierdzenie Graniczne – rozkłady α -stabilne (Lévi-stabilne).

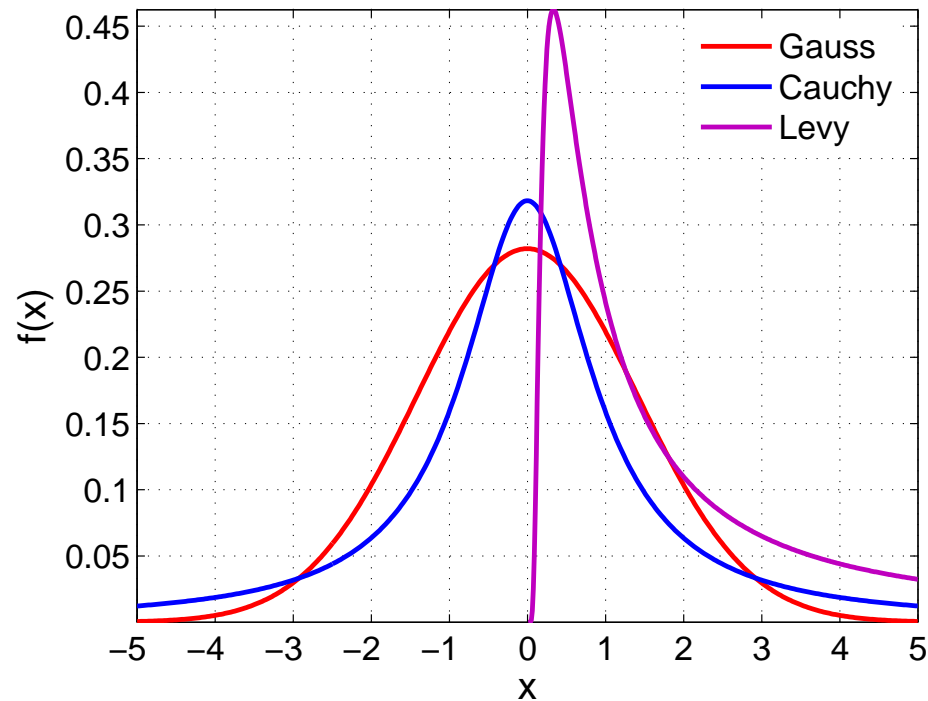
Rozkłady α -stabilne – funkcja charakterystyczna

$$X \stackrel{d}{=} S_\alpha(\beta, \sigma, \mu)$$

$$\phi(k) = \begin{cases} \exp(-\sigma^\alpha |k|^\alpha \{1 - i\beta \mathbf{sign}(k) \operatorname{tg}(-\frac{\pi\alpha}{2})\} + i\mu k) & \text{dla } \alpha \neq 1, \\ \exp(-\sigma |k| \{1 + i\beta \mathbf{sign}(k) \frac{2}{\pi} \log |k|\} + i\mu k) & \text{dla } \alpha = 1. \end{cases}$$

$0 < \alpha \leq 2$ – indeks stabilności; $\sigma > 0$ – par. skali;
 $-1 \leq \beta \leq 1$ – par. skośności.; μ – par. lokalizacji

Rozkłady Gaussa, Cauchy'ego i Lévy'ego



$$S_2(0, \sigma, \mu) = N(\mu, \sqrt{2}\sigma) \quad S_1(0, \sigma, \mu) = C(\mu, \sigma) \quad S_{\frac{1}{2}}(1, \sigma, \mu) = Levy(\mu, \sigma)$$

Suma zmiennych o rozkładzie α -stabilnym

Niech $X_1 \stackrel{d}{=} S_\alpha(\beta_1, \sigma_1, \mu_1)$ i $X_2 \stackrel{d}{=} S_\alpha(\beta_2, \sigma_2, \mu_2)$

będą zmiennymi niezależnymi, wówczas

$$X_1 + X_2 \stackrel{d}{=} S_\alpha(\beta, \sigma, \mu)$$

gdzie

$$\beta = \frac{\beta_1 \sigma_1^\alpha + \beta_2 \sigma_2^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha}, \quad \sigma^\alpha = \sigma_1^\alpha + \sigma_2^\alpha, \quad \mu = \mu_1 + \mu_2.$$

Wybrane własności rozkładów α -stabilnych

- ✗ Istnienie momentów dla $\alpha < 2$

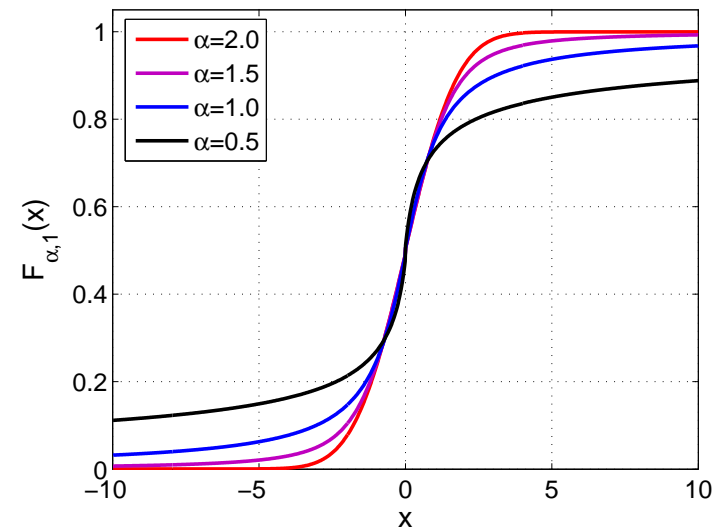
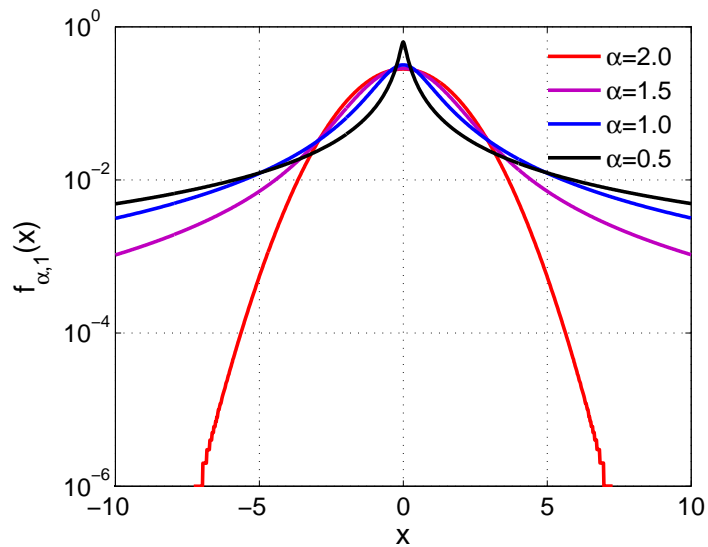
$$E(X^p) = \int_{-\infty}^{\infty} x^p f(x) dx < +\infty \Leftrightarrow 0 < p < \alpha.$$

- ✗ Istnienie algorytmów generujących $X \stackrel{d}{=} S_\alpha(\beta, \sigma, \mu)$

- ✗ Symetryczne rozkłady α -stabilne

$$S_\alpha S(\sigma) = S_\alpha(0, \sigma, 0); \phi(k) = \exp(-\sigma^\alpha |k|^\alpha).$$

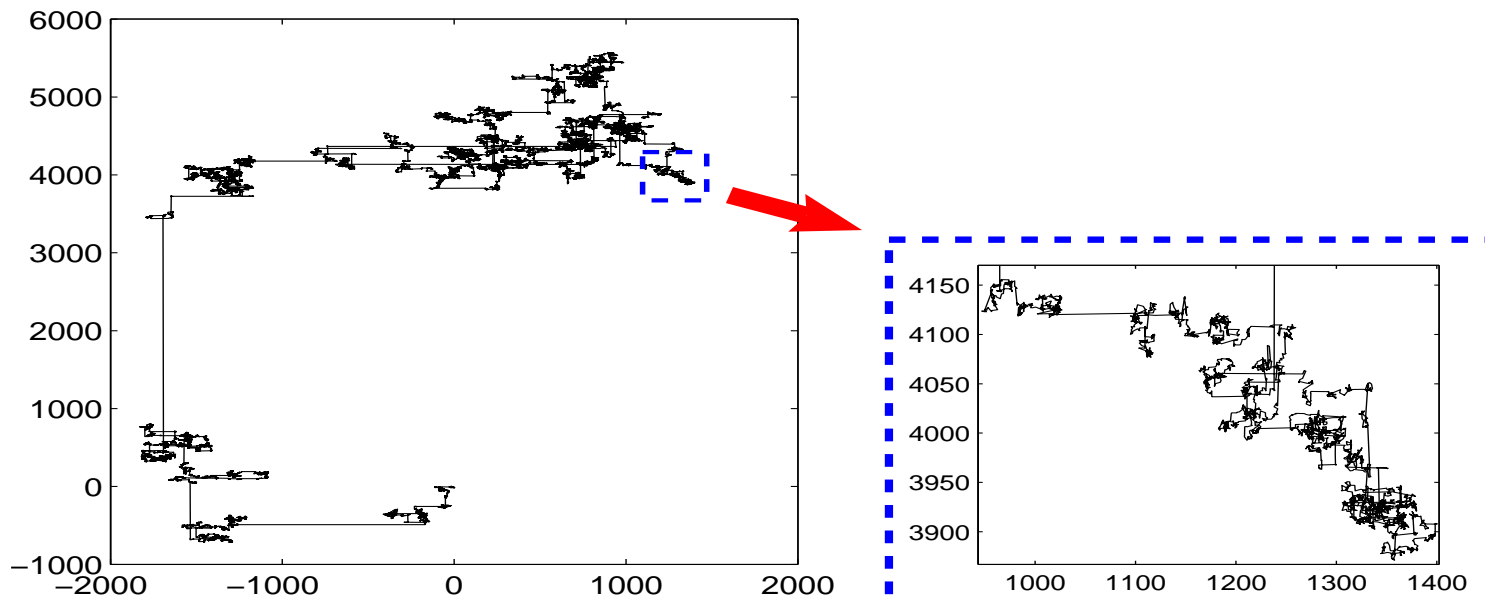
Symetryczne rozkłady α -stabilne



Funkcje gęstości $f_{\alpha,1}(x)$ i dystrybuanty $F_{\alpha,1}(x)$ zmiennej losowej $S\alpha S(1)$

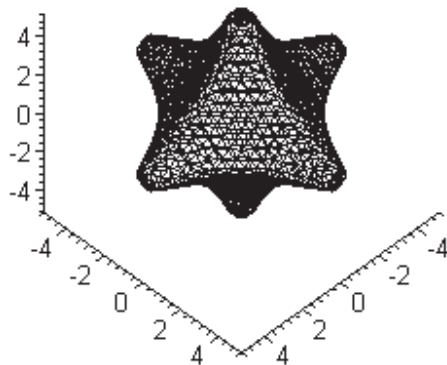
Mutacja w EA oparta na $S\alpha S(\sigma)$

$$x'_i = x_i + S\alpha S(\sigma), \quad (x' = x + S\alpha S(\sigma))$$

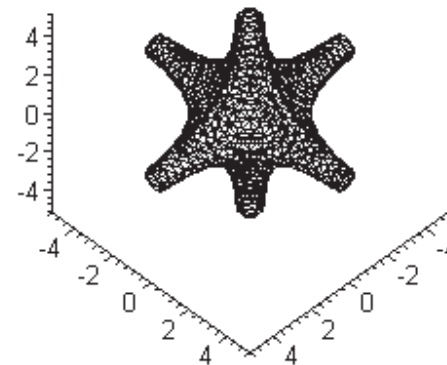


Nieizotropowość $S_{\alpha}S(\sigma)$

(a)



(b)



Izopowierzchnie funkcji gęstości 3D ($f_{\alpha}(x) = 0.001$) rozkładów $S_{\alpha}S$ dla $\alpha = 1$ (a), and $\alpha = 0.5$ (b).

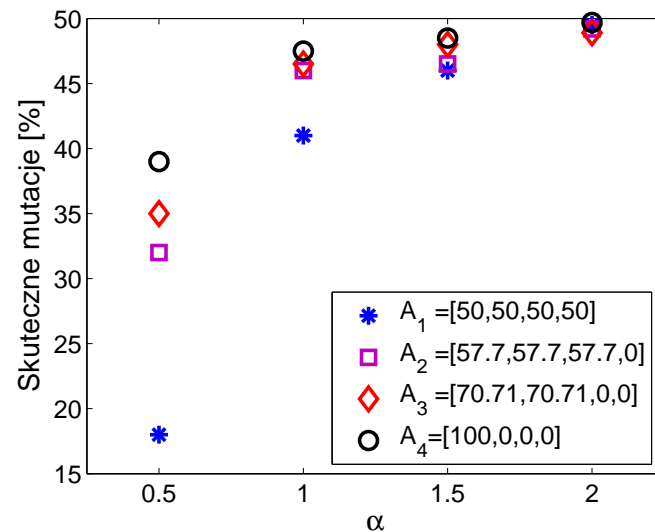
Nieizotropowość $S_\alpha S(\sigma)$ – konsekwencje

Eksperyment

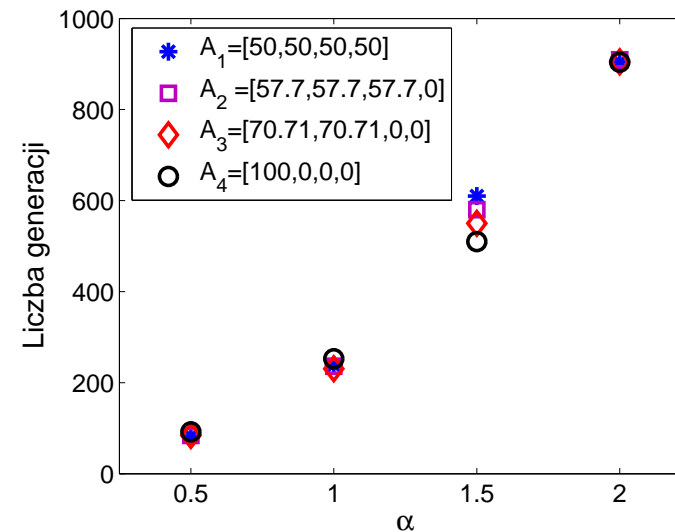
- ✓ funkcja dopasowania – $\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$, $\mathbf{x} \in \mathbb{R}^4$
(funkcja sferyczna);
- ✓ algorytmy – $(1 + 1)ES_\alpha$, $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2$, $\sigma = 0.1$;
- ✓ punkty startowe: $\mathbf{a}_1 = (100, 0, 0, 0)$, $\mathbf{a}_2 = (\frac{100}{\sqrt{2}}, \frac{100}{\sqrt{2}}, 0, 0)$,
 $\mathbf{a}_3 = (\frac{100}{\sqrt{3}}, \frac{100}{\sqrt{3}}, \frac{100}{\sqrt{3}}, 0)$, $\mathbf{a}_4 = (50, 50, 50, 50)$.

Nieizotropowość $S_\alpha S(\sigma)$ – konsekwencje

(a)



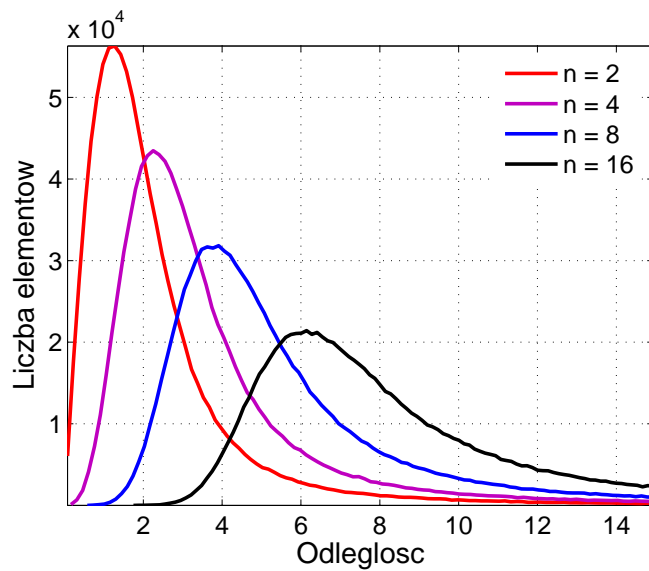
(b)



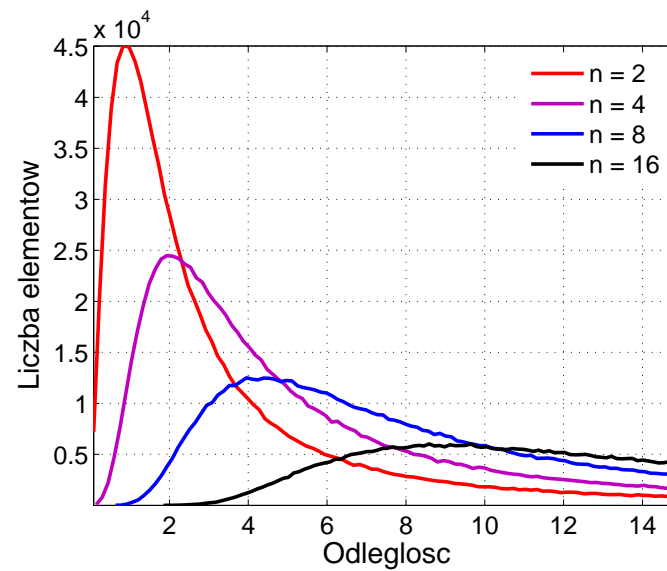
Procent skutecznych mutacji (a) i średnia liczba iteracji konieczna do lokalizacji optimum ($\Phi(x) < 0.5$) (b) vs. α

Efekt martwego otoczenia – rozkłady $\chi_{\alpha,n}$

$$\chi_{\alpha,n} = \sqrt{\sum_{k=1}^n X_k^2}, \quad X_k \stackrel{d}{=} S\alpha S(\sigma),$$



$$\alpha = 1.5$$



$$\alpha = 1$$

Histogramy długości 10^6 wektorów losowych $\mathbf{X} \stackrel{d}{=} S\alpha S(1)$

Efekt martwego otoczenia – konsekwencje

Eksperyment

✓ funkcja dopasowania

✓ $\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$ (f. sferyczna);

✓ $\Phi_r(\mathbf{x}) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$

(uogólniona f. Rosenbrocka)

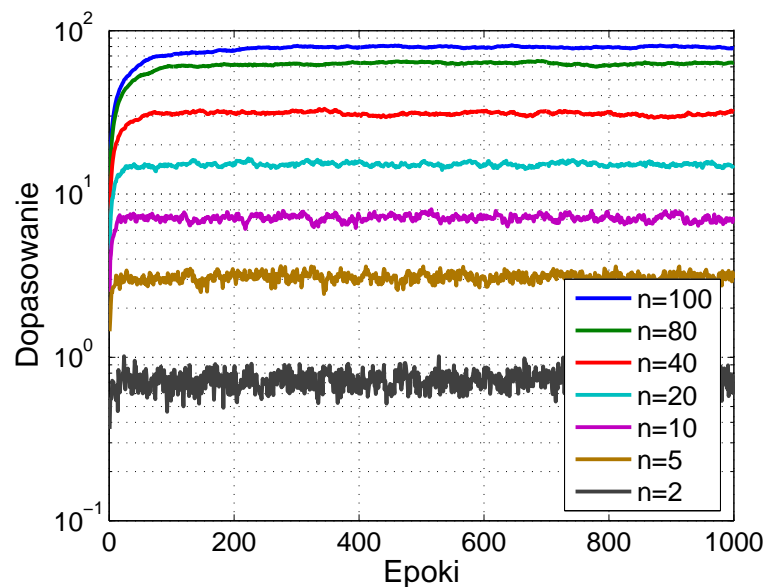
✓ algorytmy – $ESSS_\alpha$, $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2$, $\sigma = 0.05$;

✓ punkty startowe: $\mathbf{x}_0 = (0, 0, \dots, 0)$ dla $\Phi_s(\mathbf{x})$,

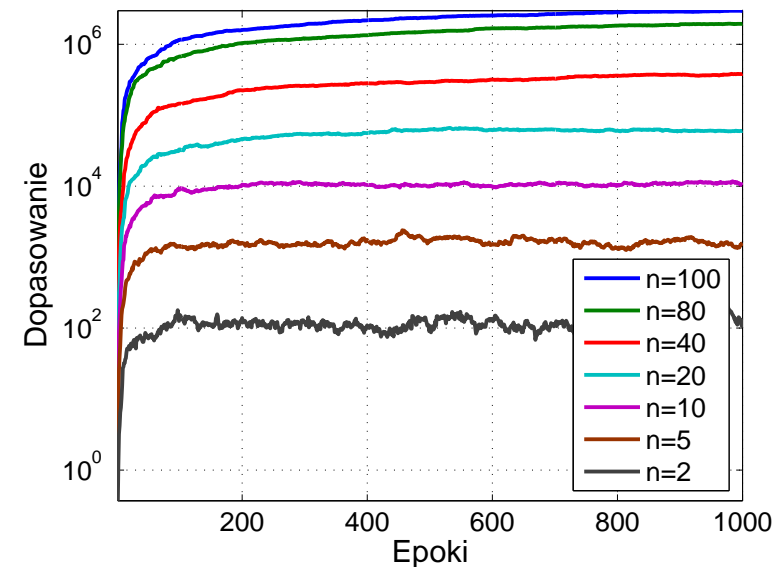
$\mathbf{x}_0 = (\frac{30}{\sqrt{n}}, \dots, \frac{30}{\sqrt{n}})$, dla $\Phi_r(\mathbf{x})$.

Efekt martwego otoczenia – konsekwencje

(a)



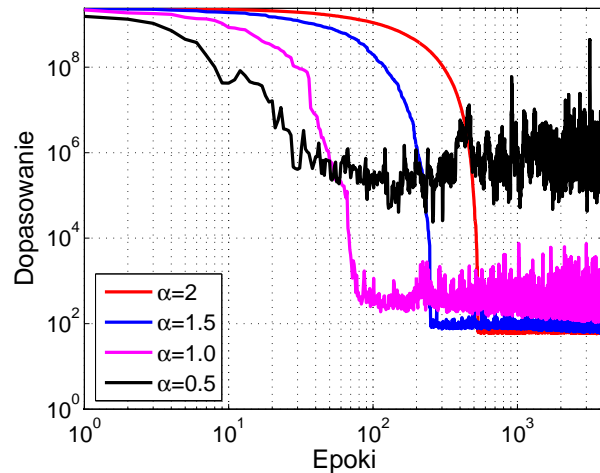
(b)



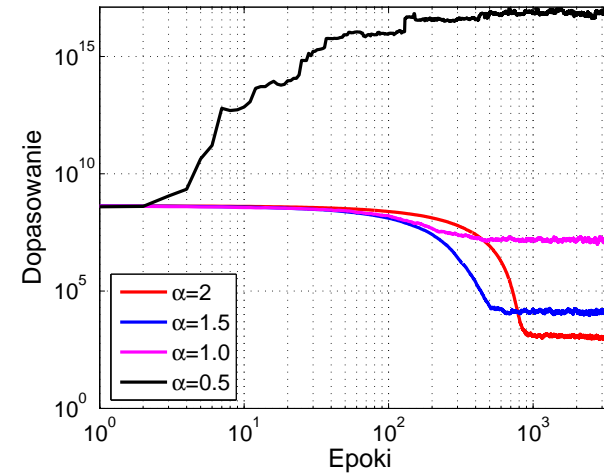
Odległość najlepszego osobnika populacji w $ESSS_2$ (a) i $ESSS_{\frac{1}{2}}$ od optimum $\Phi_s(\mathbf{x})$ vs. iteracje (średnie po 100 uruchomieniach).
 ($n = 2, 5, 10, 20, 40, 60, 80, 100$ linie od dołu do góry.)

Efekt martwego otoczenia – konsekwencje

(a)



(b)



Najlepsze dopasowanie $\Phi_r(\mathbf{x})$ w populacji $ESSS_\alpha$ vs. iteracje (średnie po 500 uruchomieniach) w środowisku 2D (a) i 20D (b).

($\alpha = 2$ – l. ciągła, $\alpha = 1.5$ – l. kres., $\alpha = 1$ – l. kres.-krop., $\alpha = 0.5$ – l. krop.)

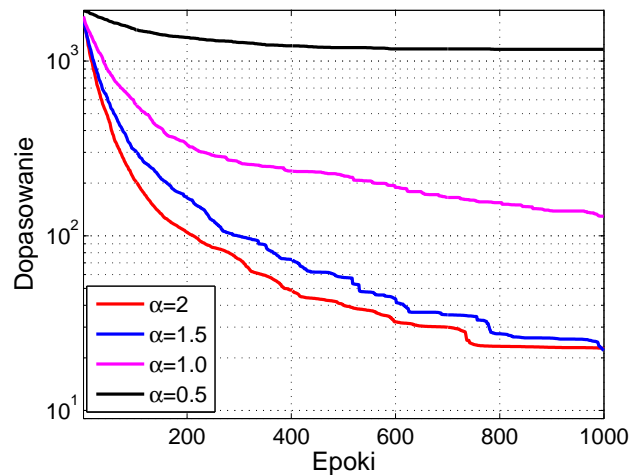
Efekt martwego otoczenia + adaptacja parametru skali

Eksperyment

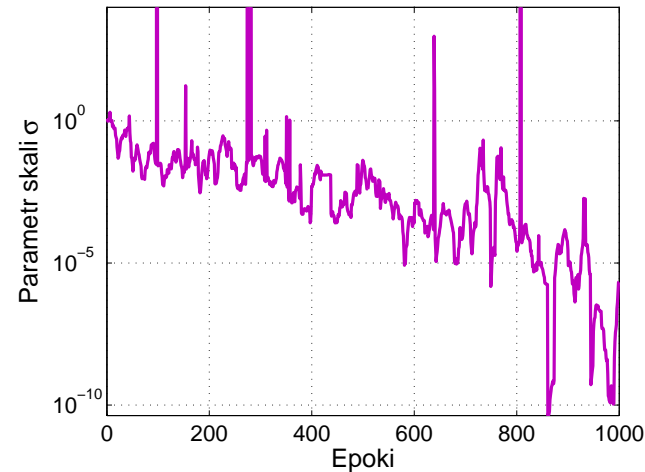
- ✓ funkcja dopasowania – $\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$;
- ✓ algorytmy – EP_α , $\alpha = \frac{1}{2}, 1, \frac{3}{2}, 2$;
- ✓ rozmiar populacji $\eta = 20$, rozmiar sparingu $q = 5$;
- ✓ obszary inicjacji: $\Omega_x = [0, 8, 1, 2] \times \prod_{i=2}^n [-0, 2, 0, 2]$,
 $\Omega_\sigma = \prod_{i=1}^n [0, 0.01]$.

Efekt martwego otoczenia + adaptacja parametru skali

(a)



(b)



(a) Najlepsze dopasowanie $\Phi_s(\mathbf{x})$ w populacji EP_α vs. iteracje (średnie po 500 uruchomieniach) w środowisku 20D.

($\alpha = 2$ – l. ciągła, $\alpha = 1.5$ – l. kres., $\alpha = 1$ – l. kres.-krop., $\alpha = 0.5$ – l. krop.)

(b) średnia wartość σ w populacji dla EP_1 .

Izotropowa mutacja α -stabilna

$$\mathbf{x}' = \mathbf{x} + \mathbf{X} = \mathbf{x} + r\mathbf{u}^{(n)}$$

gdzie $\mathbf{u}^n \stackrel{d}{=} U(\text{Sphere}(1))$; $r \stackrel{d}{=} |S\alpha S(\sigma)|$;

Funkcja gęstości $\mathbf{X} \stackrel{d}{=} \text{IS}\alpha\text{S}(\sigma)$:

$$g(\mathbf{x} \mid \alpha, \sigma, \boldsymbol{\mu}, n) = \frac{1}{\sigma \pi^{n/2}} \frac{\Gamma(\frac{n}{2})}{\|\mathbf{x} - \boldsymbol{\mu}\|^{n-1}} f_{\alpha,1} \left(\frac{\|\mathbf{x} - \boldsymbol{\mu}\|}{\sigma} \right)$$

Wskaźnik progresu dla $(1 + 1)ES_\alpha$

Założenia:

- ✓ funkcja dopasowania – $\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$,
- ✓ mutacja elementu \mathbf{y}_k : $\mathbf{y}'_k = \mathbf{y}_k + |S\alpha S(\sigma)|\mathbf{u}$,
- ✓ twarda selekcja: $\mathbf{y}_{k+1} = \arg \min\{\Phi_s(\mathbf{y}_k), \Phi_s(\mathbf{y}'_k)\}$.

Wskaźnik progresu dla $(1 + 1)ES_\alpha$ – formalizm

Niech $V = \frac{\|\mathbf{Y}'_k\|^2}{\|\mathbf{y}_k\|^2}$ wówczas

$$\varphi = E \{ \min\{V, 1\} \mid \|\mathbf{y}_k\| \} = 1 - \int_0^1 (1 - v) f_v(v) dv,$$

gdzie

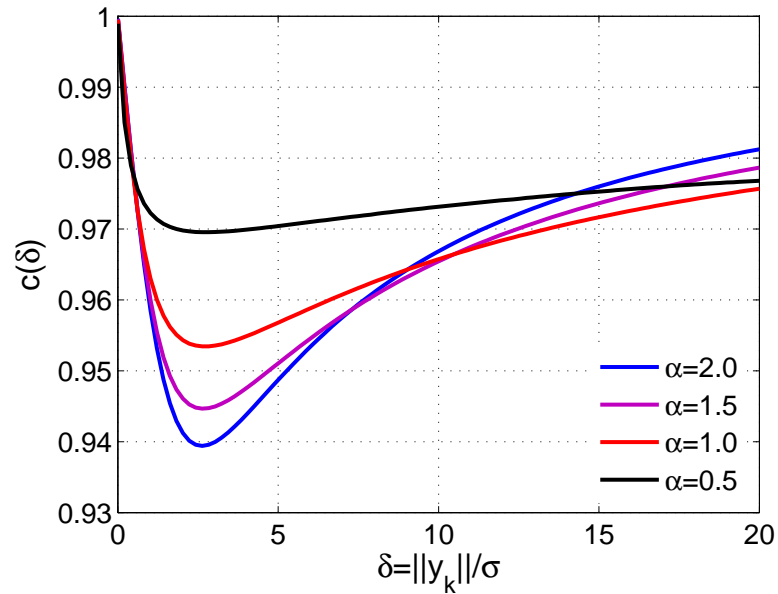
$$\color{red}{/} f_v(v) = \frac{\delta v^{\frac{n}{2}-1}}{B\left(\frac{n-1}{2}, \frac{1}{2}\right)} \int_{-1}^1 \frac{f_{\alpha,1}\left(\delta(v-2t\sqrt{v+1})^{\frac{1}{2}}\right)}{(v-2t\sqrt{v+1})^{\frac{(n-1)}{2}}} (1-t^2)^{\frac{(n-3)}{2}} dt -$$

funkcja gęstości V ;

$$\color{red}{/} \delta = \frac{\|\mathbf{y}_k\|}{\sigma} - \text{parametr odchyłu od centrum};$$

$$\color{red}{/} B(\cdot, \cdot) - \text{funkcja Beta.}$$

Wskaźnik progresu dla $(1 + 1)ES_\alpha$ – wyniki



**Wskaźnik progresu dla $(1 + 1)ES_\alpha$ vs. $\delta = \|y_k\|/\sigma$.
Wartości dokładne dla $\alpha = 0.5, 1, 1.5, 2.0$.**

Wskaźnik progresu dla $(1 + \lambda)ES_\alpha$

Założenia:

✓ funkcja dopasowania – $\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$,

✓ mutacja \mathbf{y}_k :

$$\mathbf{y}'_{k,i} = \mathbf{y}_k + \mathbf{X}_i, \quad \mathbf{X}_i \sim |S\alpha S(\sigma)| \frac{N(0, \mathbf{I})}{\|N(0, \mathbf{I})\|}, \quad i = 1, 2, \dots, \lambda,$$

✓ twarda selekja:

$$\mathbf{y}_{k+1} = \arg \min \{ f(\mathbf{y}_k), \{ f(\mathbf{y}'_{k,i}) \mid i = 1, 2, \dots, \lambda \} \}.$$

Wskaźnik progresu dla $(1 + \lambda)ES_\alpha$ – formalizm

Niech

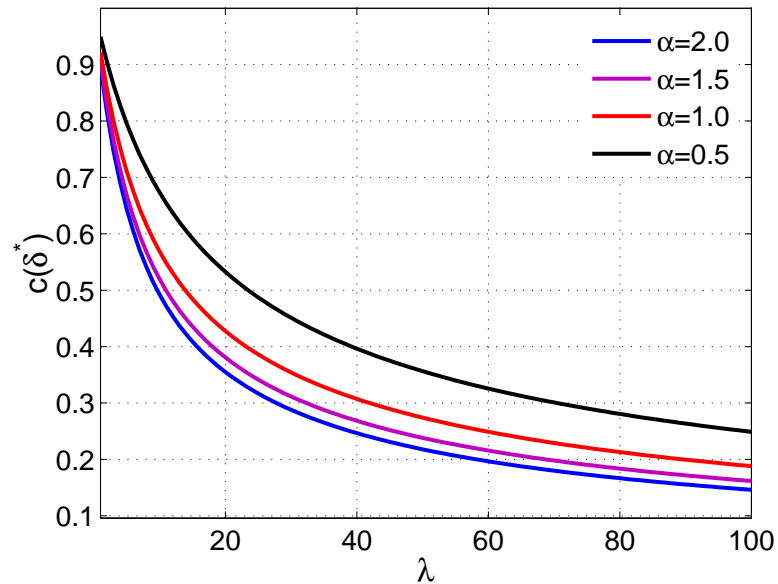
$(\mathbf{X}_i | i = 1, \dots, \lambda) \rightarrow (\mathbf{X}_{i:\lambda} | i = 1, \dots, \lambda) : (i < j \Rightarrow \|\mathbf{X}\|_{i:\lambda} \leq \|\mathbf{X}\|_{j:\lambda})$.

Let $V_{i:\lambda} = \frac{\|\mathbf{Y}'_{k,i}\|^2}{\|\mathbf{y}_k\|^2}$ **wówczas**

$$\varphi = E \{ \min\{V_{1:\lambda}, 1\} \mid \|\mathbf{y}_k\| \} = \int_0^1 (1 - F_v(v))^\lambda dv,$$

gdzie $F_v(v)$ **jest** dystrybuantą $V_{i:\lambda}$.

Wskaźnik progresu dla $(1 + \lambda)ES_\alpha$ – wyniki



Maksymalny wskaźnik progresu $(1 + \lambda)ES_\alpha$ ($\alpha = 0.5, 1, 1.5, 2$ linie od góry do dołu) vs. liczba potomków λ .

Izotropowa mutacja α -stabilna + miękka selekcja

Założenia intuicyjne

- ✓ mutacja pozwala na generowanie potomków blisko rodzica → lepsza zdolność eksploatacyjna EA;
- ✓ mutacja pozwala na częste generowanie potomków odległych → lepsza zdolność eksploracyjna EA.

Izotropowa mutacja α -stabilna + miękka selekcja

Wspólne cechy EA z miękka selekcją

- ✓ populacja w iteracji $k + 1$ jest zwykle utworzona na bazie pewnego podzbioru k -tej populacji;
- ✓ \bar{r} – wartość oczekiwana liczby potomków w generacji $k + 1$ posiadającego tego samego rodzica;
- ✓ analiza lokalnej zbieżności jest zredukowana do analizy wartości oczekiwanej pierwszej zmiennej $X_{1:[\bar{r}]}$ statystyki porządkowej.

Istnienie k-tego momentu $X_{1:\lambda}$

Twierdzenie:

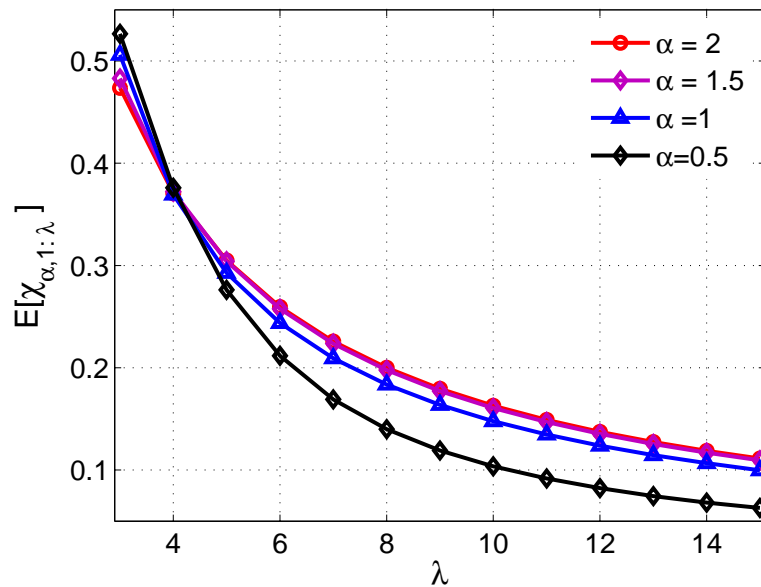
Niech $X \sim |S_{\alpha S}|$ wówczas

$$k - \alpha(\lambda - i + 1) < 0 \Rightarrow E\{X_{i,\lambda}^k\} < +\infty$$

Wniosek

 $\lambda > 1/\alpha \Rightarrow E\{X_{1,\lambda}\} < +\infty;$

Izotropowa mutacja α -stabilna + miękka selekcja



Wartość oczekiwana zmiennej losowej $X_{1:\lambda}$ vs. λ .

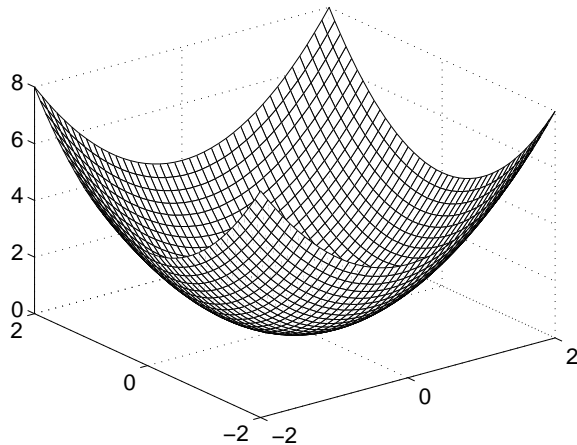
$X \stackrel{d}{=} |S_2 S(1)|$ – kwadraty,

$X \stackrel{d}{=} |S_{1.5} S(1)|$ – diamenty,

$X \stackrel{d}{=} |S_1 S(1)|$ – krzyże,

$X \stackrel{d}{=} |S_{0.5} S(1)|$ – koła

Estymacja rozmiaru martwego obszaru



Funkcja dopasowania:

$$\Phi_s(\mathbf{x}) = \|\mathbf{x}\|^2$$

Populacja początkowa:

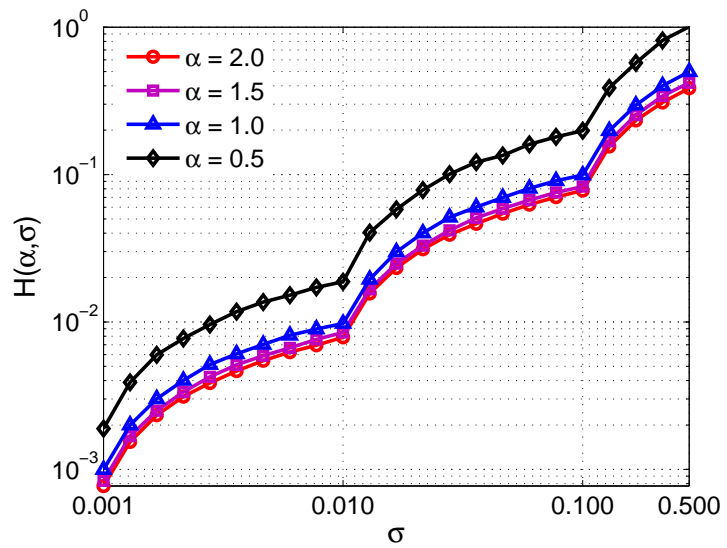
$$P_0 = \{\mathbf{x}_i^0\}_{i=1}^{\eta}, \mathbf{x}_i^0 = \mathbf{0}$$

Algorytm:

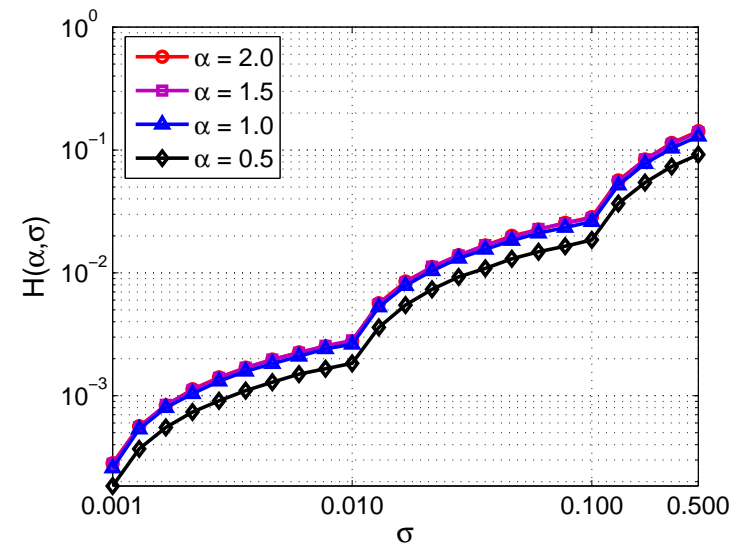
selekcja turniejowa + IS α S

**Wskaźnik jakości: $H(\alpha, \sigma)$ - średnia odległość
najlepszego osobnika od rozwiązania.**

Zdolności eksploatacyjne

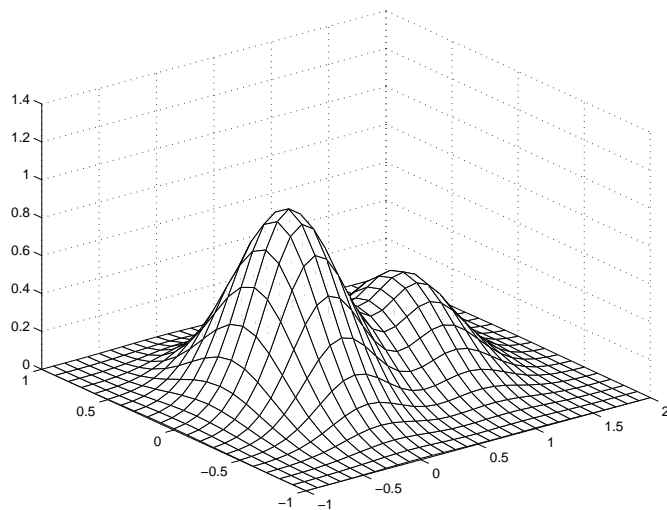


$$\lambda = 2$$



$$\lambda = 8$$

Przekraczanie siodła - właściwości eksploracyjne



Populacja początkowa:

$$P_0 = \{\mathbf{x}_i^0\}_{i=1}^{\eta},$$

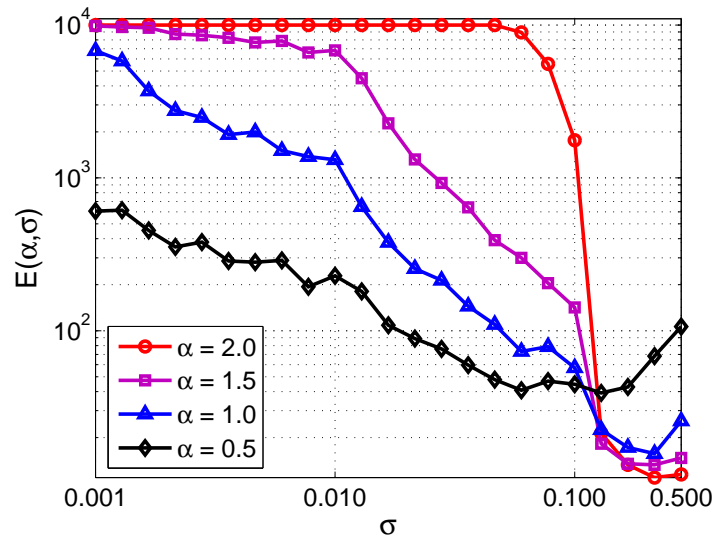
\mathbf{x}_i^0 – optimum lokalne

Warunek stopu:

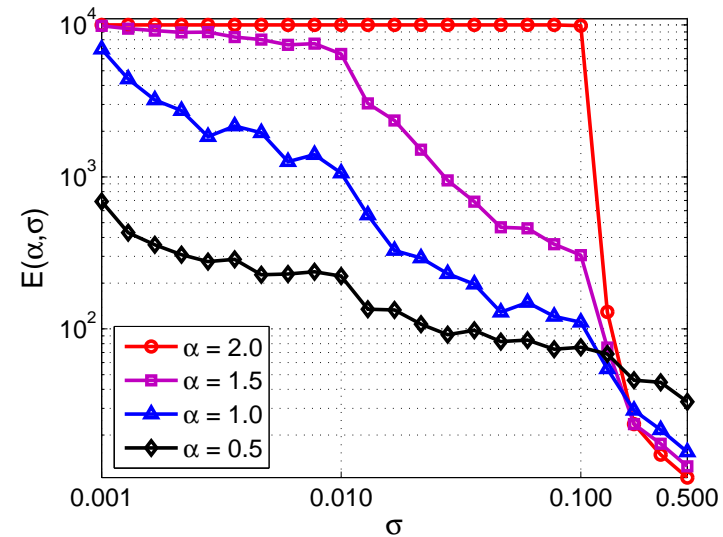
$$\max_{\mathbf{x}_i^k \in P(k)} \{\phi(\mathbf{x}_i^k)\} > 0.6$$

Wskaźnik jakości: $E(\alpha, \sigma)$ - średnia liczba iteracji potrzebna do rozwiązania problemu.

Zdolności eksploracyjne - przekraczanie siodła

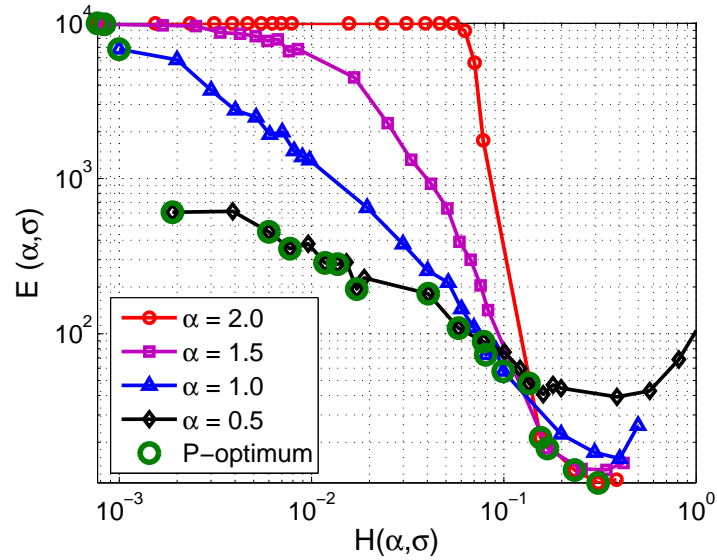


$$\lambda = 2$$

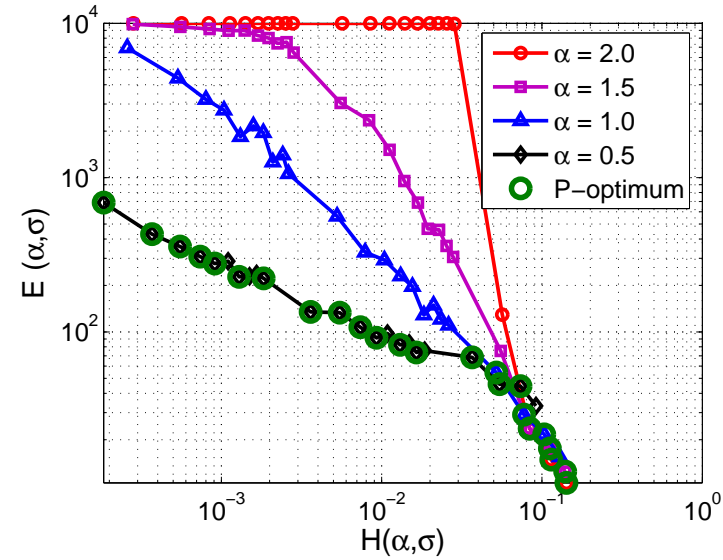


$$\lambda = 8$$

Eksploracja vs. Eksploatacja – problem dwukryterialny



$$\lambda = 2$$



$$\lambda = 8$$

Podsumowanie

- ✘** Otrzymane dotychczasowe wyniki badań podważają dotychczasową hegemonię mutacji gaussowskiej w stochastycznych algorytmach optymalizacji globalnej.
- ✘** Istnieje możliwość pogodzenia wymagań eksploatacyjnych i eksploracyjnych algorytmów ewolucyjnych stosując izotropową mutację α -stabilną.

Aktualne realizowane zagadnienia

- ✗ samoadaptacja parametrów mutacji α -stabilnej:
 - + parametr skali σ ;
 - + indeksu stabilności α ;
- ✗ wymuszony kierunek mutacji;
 - + samoadaptacja kierunku wymuszenia;
 - + zastosowanie w zagadnieniach hyper-multiwymiarowych i środowiskach niestacjonarnych;
- ✗ lokalna selekcja + izotropowa mutacja α -stabilna;
 - + mutacja promienia selekcji;
 - + zastosowanie w zagadnieniach wielokryterianych i środowiskach niestacjonarnych;

Ważniejsze publikacje

✗ Rozkłady Cauchy'ego w EA:

- ✚ G. Rudolph, Local convergence rates of simple evolutionary algorithms with Cauchy mutations – *IEEE TEC*, Vol.1, No.4, pp.249–258, 1997.
- ✚ X. Yao, Y. Liu, Fast evolutionary strategies – *Contr. Cybern.*, Vol.26, No.3, pp.467–496, 1997;
- ✚ X. Yao, Y. Liu, Evolutionary programming made faster – *IEEE TEC*, Vol.3, No.2, pp.82–102, 1999;
- ✚ A. Obuchowicz, Multidimensional mutations in evolutionary algorithms based on real-valued representation, *Int. J. System Science*, Vol.34, No.7, 2003, pp.469–483.

× Nieizotropowe mutacje α -stabilne w EA;

+ A. Obuchowicz, P. Prętki, Phenotypic Evolution with Mutation Based on Symmetric α -Stable Distributions. – *Int. J. Appl. Math. and Comp. Sci.* Vol.14, No.3, pp.289-316, 2004;

+ C.Y. Lee, X. Yao, Evolutionary programming using mutations based on the Lévy probability distribution. – *IEEE TEC* Vol.8, No.1, pp.1–13, 2004;

× Izotropowe mutacje α -stabilne w EA;

+ A. Obuchowicz, P. Prętki, Isotropic Symmetric α -Stable Mutations for Evolutionary Algorithms. *Proc. IEEE CEC'05*, Vol. 1, pp. 404-410, 2005;