Prednáška 3

Optimalizačné metódy pre funkcie
n-premenných

Študujme **reálnu funkciu** n-premenných

$$f: R^n \to R$$

Našou úlohou bude nájsť také $x_{\text{opt}} \in R^{\text{n}}$, pre ktoré má funkcia f minimum

$$\boldsymbol{x}_{opt} = \arg\min f(\boldsymbol{x})$$

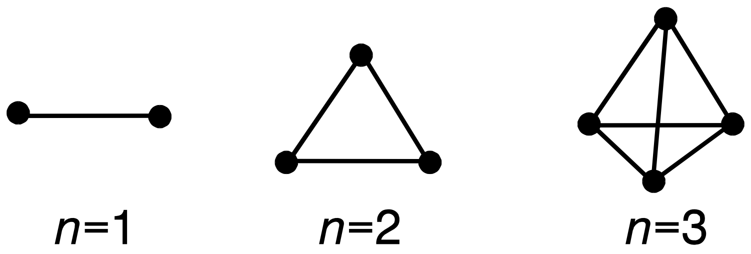
Túto úlohu budeme riešiť pomocou:

(1) gradientových metód
(2) negradientových metód

Typickým predstaviteľom negradientových metód na hľadanie minima funkcií n-premen-ných je simplexova metóda.

1. Simplexova metóda

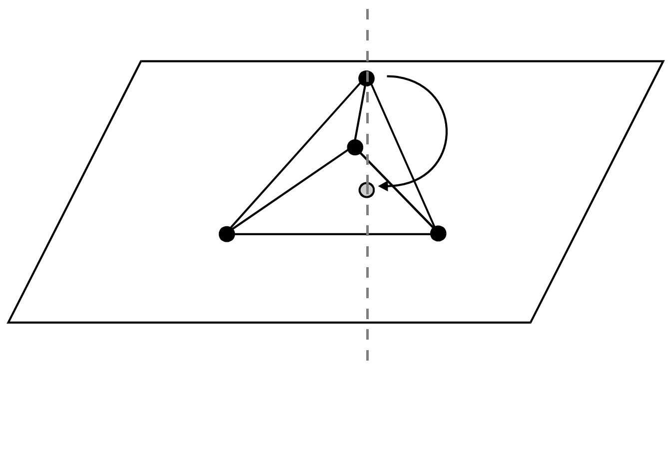
Pod **simplexom** v n-rozmernom priestore R^n rozumieme množinu n+1 bodov - vrcholov, pričom každá dvojica vrcholov je spojená hranou.



Formálne, simplex vyjadríme ako množinu

$$S = \{P_1, P_2, \ldots, P_{n+1}\} \subset R^n$$

Hovoríme, že simplex S je **nedegenerovaný**, ak žiadny z vrcholov neleží v rovine obsahujúcej ostatné vrcholy simplexu



Podmienka pre nedegenerovaný simplex môže byť preformulovaná tak, že n vektorov

$$P_2 - P_1, P_3 - P_1, \ldots, P_n - P_1, P_{n+1} - P_1,$$

je lineárne nezávislá. Nech bod P_i má tento tvar

$$P_i = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)} \right)$$

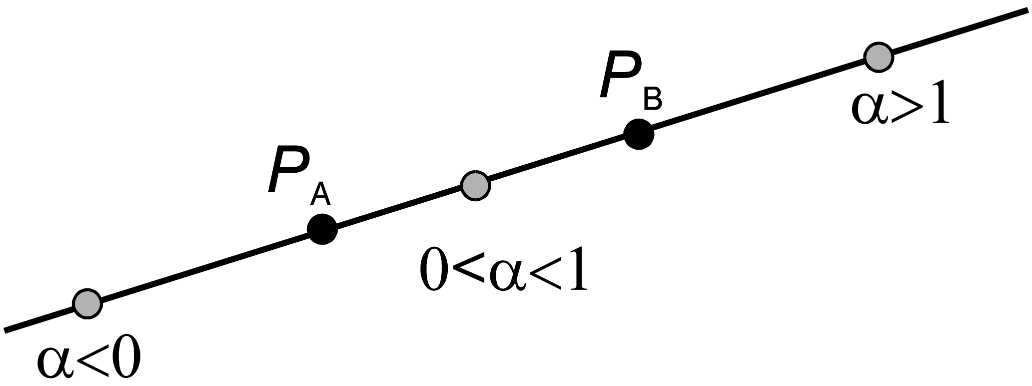
potom podmienka lineárnej nezávislosti môže byť vyjadrená pomocou podmienky nenulovosti deteminantu

$$\begin{vmatrix} x_1^{(2)} - x_1^{(1)} & x_2^{(2)} - x_2^{(1)} & \ldots\ldots. & x_n^{(2)} - x_n^{(1)} \\ x_1^{(3)} - x_1^{(1)} & x_2^{(3)} - x_2^{(1)} & \ldots\ldots. & x_n^{(3)} - x_n^{(1)} \\ \ldots\ldots. & \ldots\ldots. & \ldots\ldots. & \ldots\ldots. \\ x_1^{(n+1)} - x_1^{(1)} & x_2^{(n+1)} - x_2^{(1)} & \ldots\ldots. & x_n^{(n+1)} - x_n^{(1)} \end{vmatrix} \neq 0$$

Priamka v priestore je určená dvoma bodmi P_A a P_B

$$P_\alpha = (1 - \alpha) P_A + \alpha P_B$$

kde $\alpha \in R$ je parameter priamky, platí $P_0 = P_A$ a $P_1 = P_B$.



Priamka je **orientovaná** v zmysle rastúceho parametra α

$$P_\alpha \prec P_\beta \Leftrightarrow \alpha < \beta$$

Ťažisko simplexu S je určené ako "aritmetický priemer" jeho vrcholov

$$P_C = \frac{1}{n+1}\sum_{i=1}^{n+1} P_i$$

Funkčné hodnoty minimalizovanej funkcie f na vrcholoch simplexu označíme

$$F_1 = f(P_1), F_2 = f(P_2), \ldots, F_n = f(P_n), F_{n+1} = f(P_{n+1})$$

Medzi vrcholmi simplexa odlíšime dva vrcholy, vrhol P_{L} a vrchol P_{H}, v ktorých má funkcia f maximálnu resp. minimálnu hodnotu

$$P_H = \arg\max_{P \in S} f(P)$$

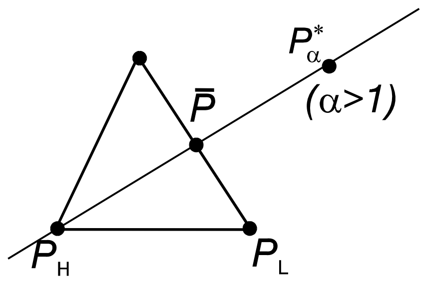
$$P_L = \arg\min_{P \in S} f(P)$$

V simplexovej metóde hrá dôležitú úlohu **ťažisko** vrcholov simplexu okrem vrcholu P_{H}

$$\overline{P} = \frac{1}{n}\left(\sum_{i=1}^{n+1} P_i - P_H\right)$$

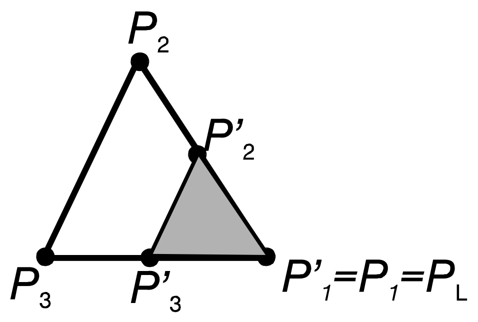
α-reflexia vrcholu P_{H} vzhľadom k ťažisku $\overline{P}$

$$P_\alpha^* = (1-\alpha)P_H + \alpha\overline{P}$$



Redukcia simplexu S vzhľadom k vrcholu P_L

$$P_i' = \frac{1}{2}\left(P_i + P_L\right) \quad (\text{pre } i = 1, 2, \ldots, n, n+1)$$



Elementárny krok simplexovej metódy

(1) Pre daný simplex S určíme vrcholy P_L a P_H

(2) Určíme ťažisko $\overline{P}$.

(3) Zostrojíme 2-reflexiu P^* vrcholu P_H okolo $\overline{P}$, $P^*=2\overline{P}-P_H$.

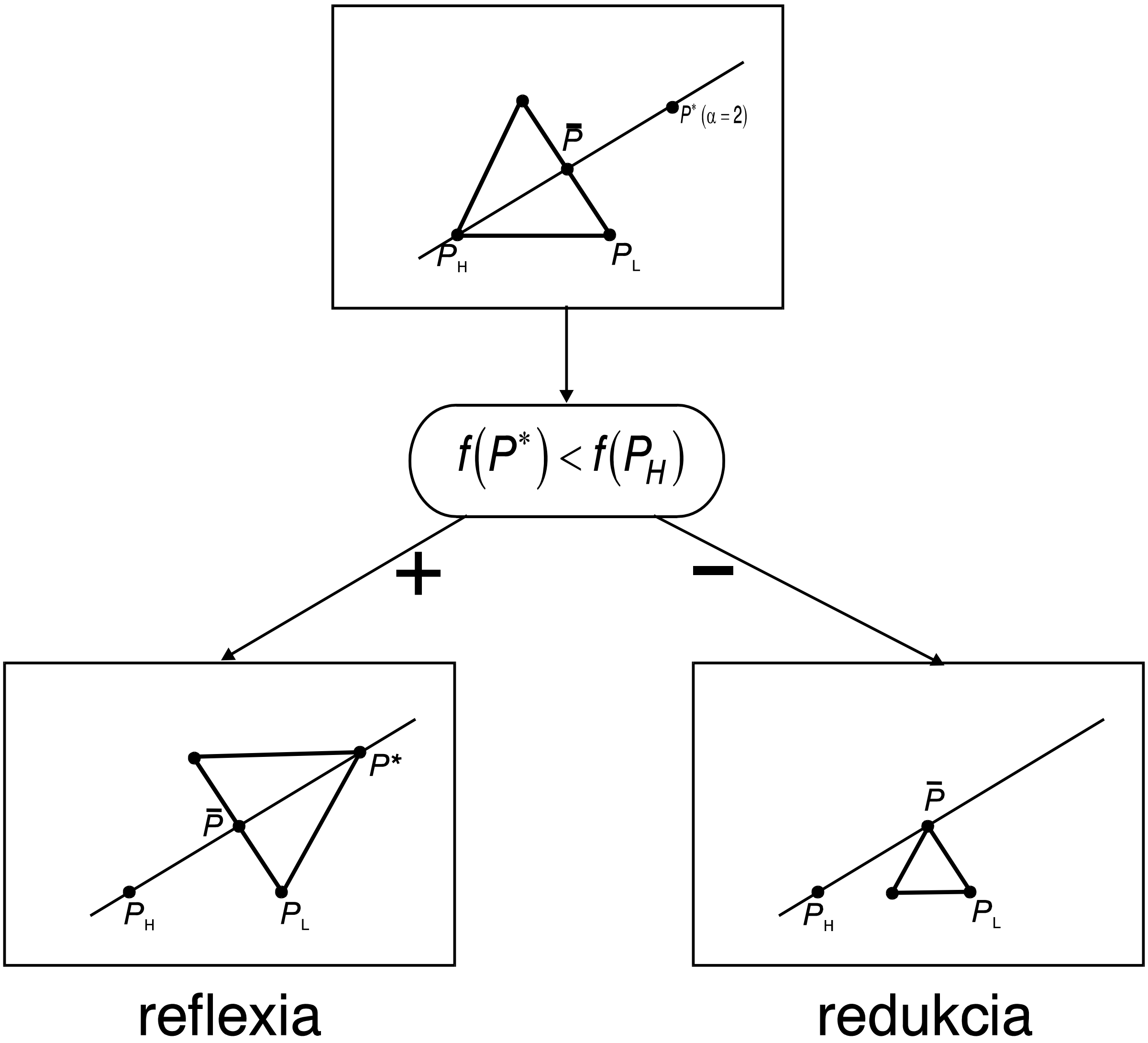
(4) Ak $f(P^*)<f(P_H)$, potom $P_H \leftarrow P^*$, v opačnom prípade vykonáme redukciu simplexa okolo vrcholu P_L.

Kroky (1-4) opakujeme tak dlho, až

$$\left|f(P_H) - f(P_L)\right| < \varepsilon$$

kde ε je malé kladné číslo (presnosť)

Diagramatická interpretácia elementárneho kroku simplexovej metódy



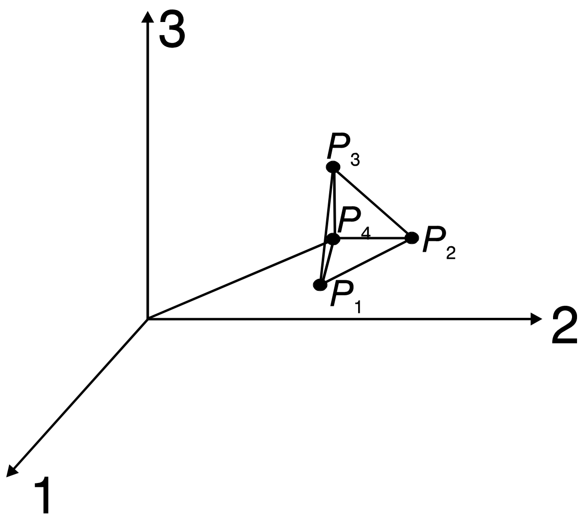
Konštrukcia počiatočného simplexu

Máme zadaný jeden vrchol P_o a dĺžku hrany d

$$P_i = P_o + d\,E_i \quad (\text{pre } i = 1,2,\ldots,n)$$
$$P_{n+1} = P_o$$

kde "body" E_i sú zadané ako jednotkové smerové vektory

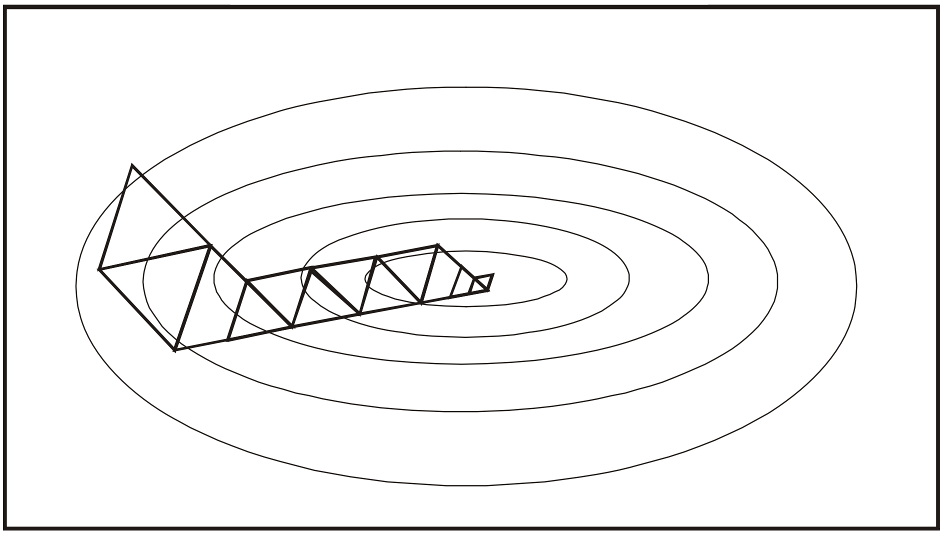
$$E_1 = (1,0,0,\ldots,0)$$
$$E_2 = (0,1,0,\ldots,0)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$E_n = (0,0,0,\ldots,1)$$



Algoritmus simplexovej metódy

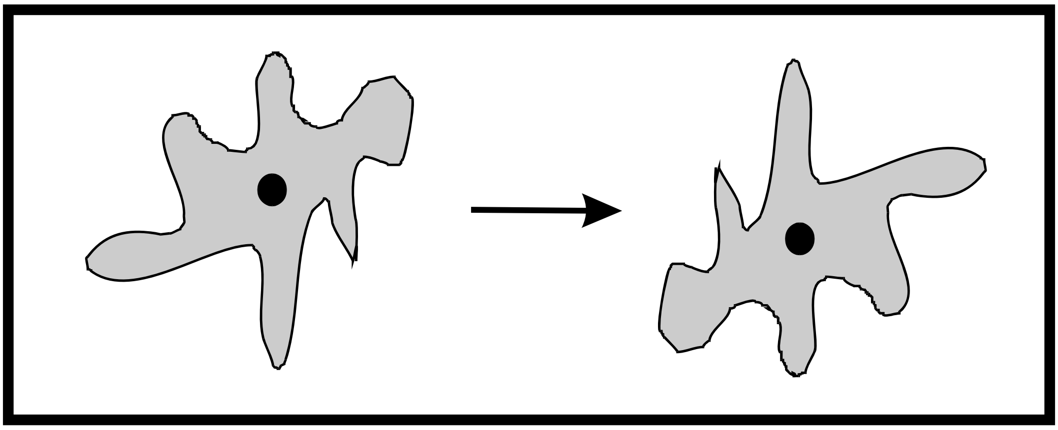
```
read(P_0,d,ε,k_max);
S:=initial set of simplex
   vertices;
norm:=∞; k:=0;
while (norm>ε) and (k<k_max) do
begin k:=k+1;
      P_H:=arg max f(P);
      P_L:=arg min f(P);
      norm:=abs(f(P_H)-f(P_L));
      P*:=2P̄-P_H;
      if f(P*)<f(P_H) then
      S:=(S-{P_H})+{P*} else
      simplex is reduced;
end;
write(P_L,f(P_L));
```

Diagramatické znázornenie priebehu simplexovej metódy



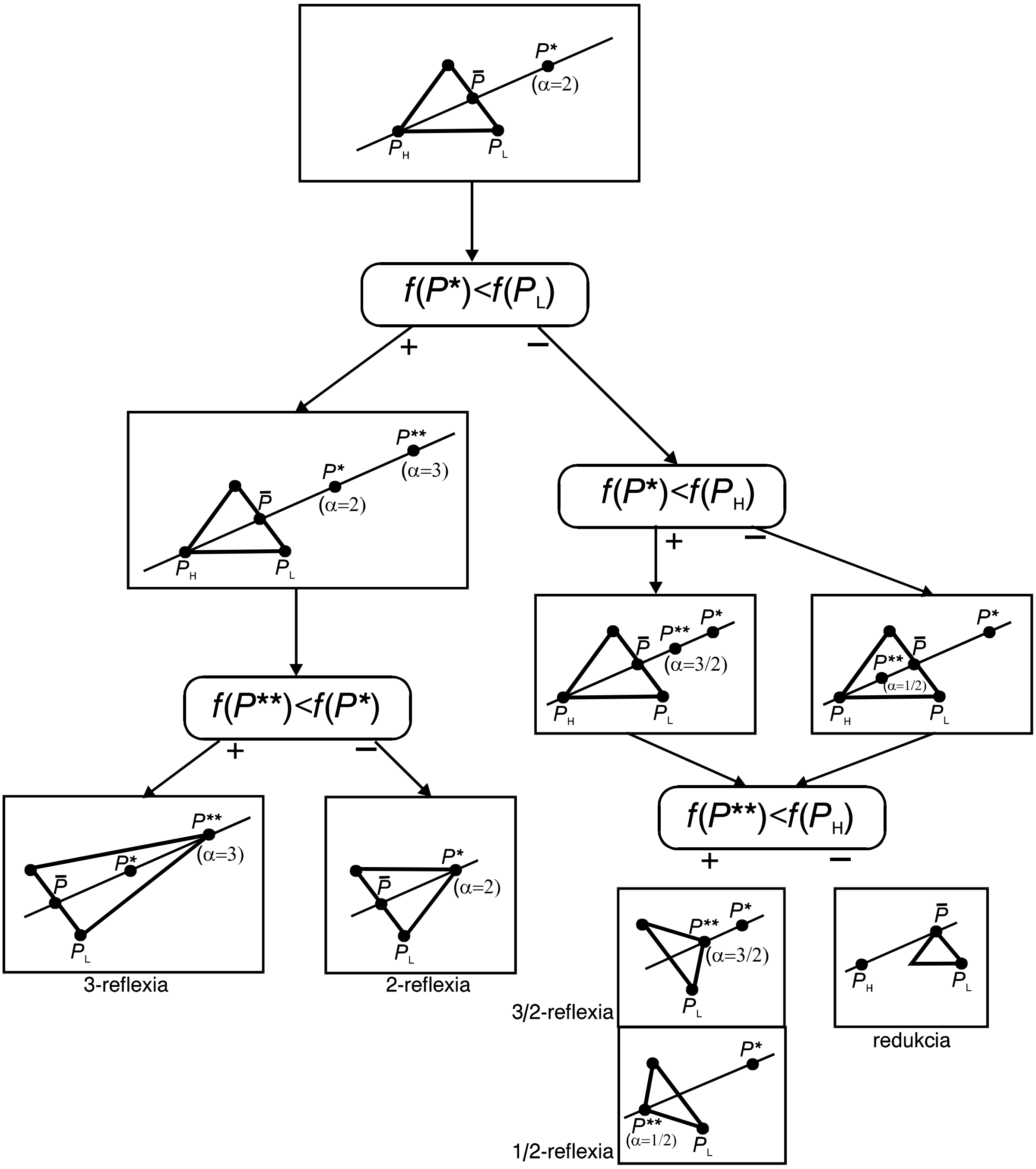
Poznámka: V priebehu simplexovej metódy dochádza často k redukcii simplexu, t.j. stáva sa stále menším a menším. Táto skutočnosť podstatne znižuje efektívnosť simplexovej metódy v blízkosti minima.

2. Modifikácia simplexovej metódy -AMÉBA



V štandardnej verzii simplexovej metódy relatívne často dochádza k redukcii simplexu, t.j. spolmaluje sa rýchlosť metódy. Tento nedostatok odstraňuje modifikácia simplexovej metódy - **améba**, v ktorej je simplex dynamicky menený v závislosti na povrchu minimalizovanej funkcie.

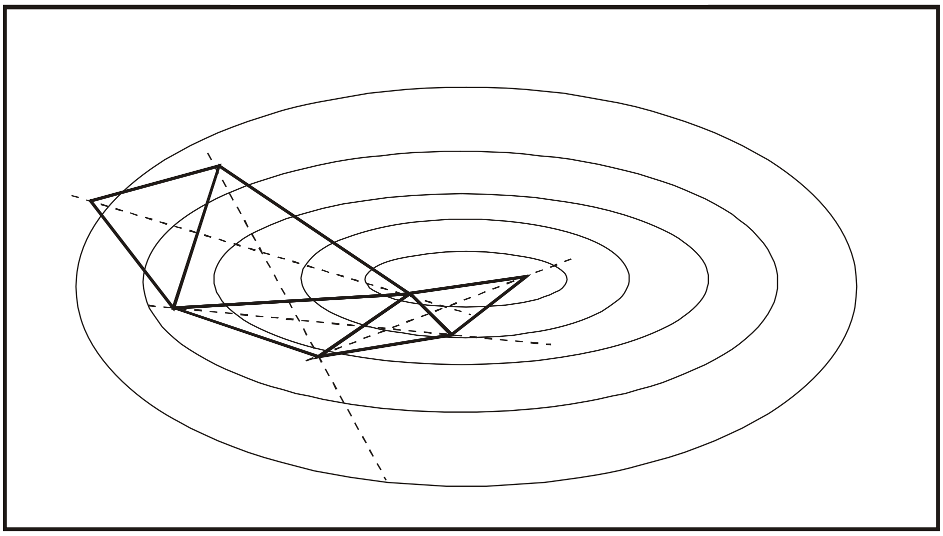
Diagramatická vizualizácia elementárneho kroku améby



Algoritmus améby

```
read(P0,d,ε,kmax);
S:=initial set of simplex
   vertices;
norm:=∞; k:=0;
while (norm>ε) and (k<kmax) do
begin k:=k+1;
      PH:=arg max f(P);
      PL:=arg min f(P);
      norm:=abs(f(PH)-f(PL));
      P*:=2P̄-PH;
      if f(P*)<f(PL) then
      begin P**:=3P̄-2PH;
            if f(P**)<f(P*) then
            S:=(S-{PH})+{P**} else
            S:=(S-{PH})+{P*};
      end else
      if f(P*)<f(PH) then
      P**:=-1/2PH+3/2P̄ else
      P**:=1/2PH+1/2P̄;
      if f(P**)<f(PH) then
      S:=(S-{PH})+{P**} else
     simplex is reduced;
end;
write(PL,f(PL));
```

Diagramatické znázornenie priebehu améby



3. Stochastická simplexová metóda

Modifikácia štandardnej simplexovej metódy sop stochastickými prvkami, ktorá je vhodná na hladanie globálneho minima.

CRS - Controlled Random Search
W.L.Price (1965)
prototyp moderných stochastických optimalizačných metód

Populácia bodov P (|P|>>n)

$$\mathrm{P} = \{P_1, P_2, \ldots, P_m\}$$

Body z populácie P s maximálnou resp. minimálnou funkčnou hodnotou sú

$$P_{max} = \arg\max_{P \in \mathrm{P}} f(P)$$

$$P_{min} = \arg\min_{P \in \mathrm{P}} f(P)$$

Simplex S obsahuje n+1 bodov ***náhodne*** vybraných z populácie P

$$S = \left\{ P_{\alpha_1}, P_{\alpha_2}, \ldots, P_{\alpha_{n+1}} \right\} \subset P$$

Štandardným postupom zostrojíme 2-reflexiu P^* v danom simplexe S .

Elementárny krok stochastickej simplexovej metódy

(1) Z populácie P náhodne výber n+1 bodov, ktoré tvoria simplex S.
(2) V rámci simplexu S zostroj 2-reflexiu P^* podobne ako v štandardnej simplexovej metóde.
(3) Ak $f(P^*)<f(P_{\max})$, potom bod $P_{\max}$ zameň za reflexiu P^* (t.j. populácia P sa modifikuje, $P \leftarrow (P \setminus \{P_{\max}\}) \cup \{P^*\}$.

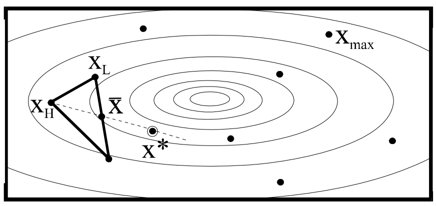
Kroky (1-3) opakuj tak dloho, až sú splnené podmienky ukončenia algoritmu.

Algoritmus stochastickej simplexovej metódy

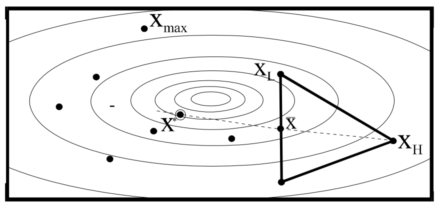
```
read(m,ε,kmax);
P:=randomly generated population
   composed of m points;
norm:=∞; k:=0;
while (norm>ε) and (k<kmax) do
begin k:=k+1;
      norm:=abs(f(Pmax)-f(Pmin));
      S:=randomly seelcted simplex
         from the population P;
      P*:= reflexion of S;
      if f(P*)<f(Pmax) then
      P:=(P\{Pmax})∪{P*};
end;
write(Pmin,f(Pmin));
```

Tri po sebe idúce elementárne kroky stochastickej simplexovej metódy

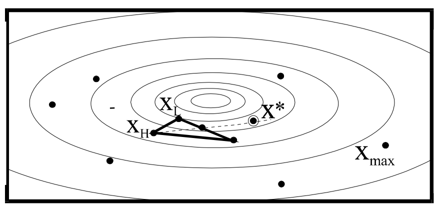
Krok 1



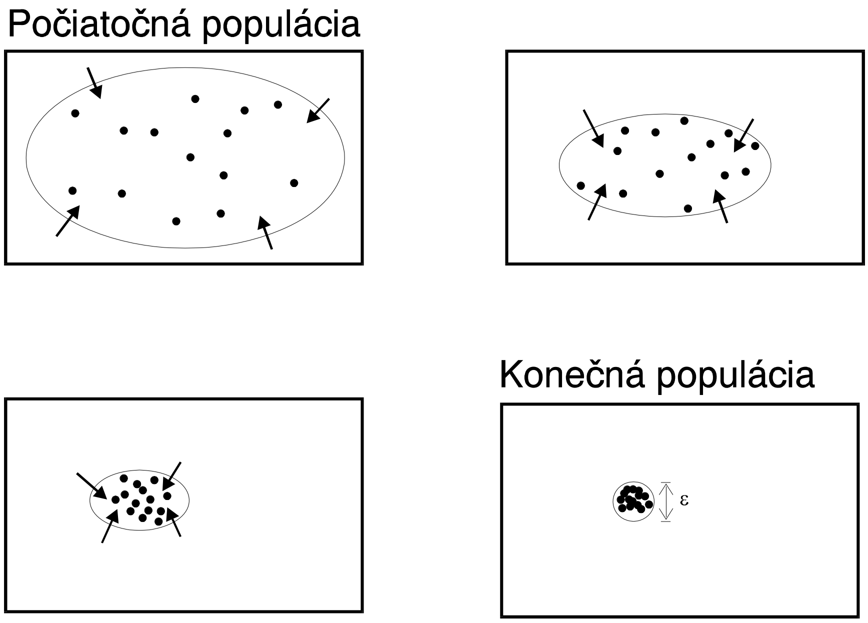
Krok 2



Krok 3



Schematické znázornenie populácií pre rôzne etapy stochastickej simplexovej metódy



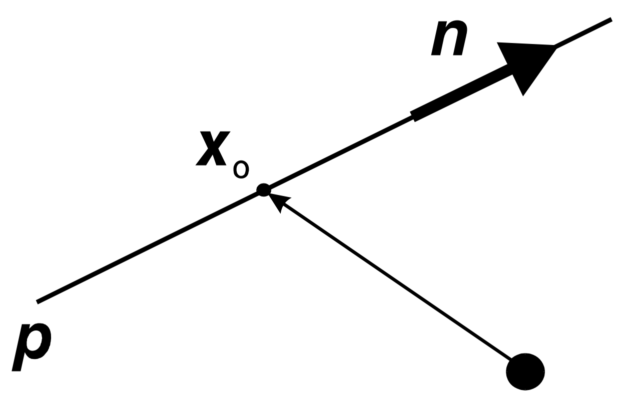
Poznámka: S rastom iteračných krokov metódy, "priemer" populácie sa zmenšuje v dôsledku toho, že sa vždy odstraňujú "okrajové" body $P_{\max}$.

4. Definícia gradientu

Študujme funkciu

$$F(\lambda) = f(\boldsymbol{x}_o + \lambda \boldsymbol{n})$$

kde $\boldsymbol{x}_\mathrm{o}=\left(x_1^o, x_2^o, \ldots, x_n^o\right) \in R^n$ je daný bod (vektor) a $\boldsymbol{n}=\left(n_1^o, n_2^o, \ldots, n_n^o\right) \in R^n$ je normalizovaný smerový vektor ($|\boldsymbol{n}|=1$). Funkcia $F(\lambda)$ popisuje "zúženie" funkcie $f(\boldsymbol{x})$ na priamku p definovanú bodom $\boldsymbol{x}_\mathrm{o}$ a smerom $\boldsymbol{n}$.



Derivácie tejto funkcie podľa premennej λ je určená vzťahom

$$F'(\lambda) = \mathrm{grad}\, f(\boldsymbol{x}_o + \lambda \boldsymbol{n}) \cdot \boldsymbol{n}$$

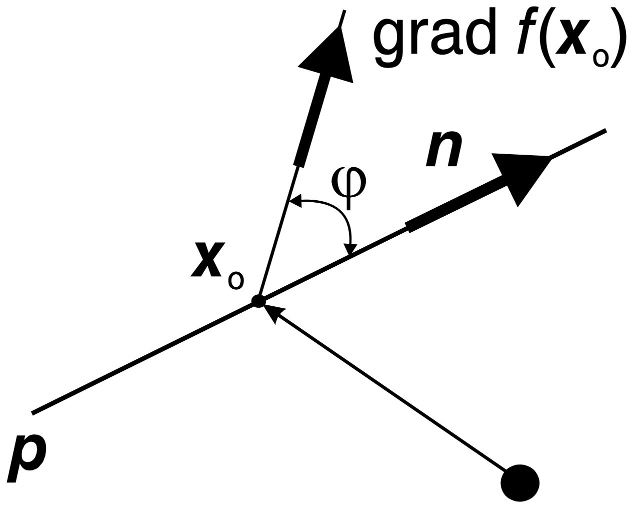
Výraz grad $f(\boldsymbol{x})$ sa nazýva **gradient** funkcie f v bode $\boldsymbol{x}$ a je definovaný ako vektor, ktorého komponenty sú 1. parciálne derivácie vzhľadom k premenným x_1, x_2,...,x_n

$$\boxed{\operatorname{grad} f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_n} \right)}$$

Hodnota derivácie F'(λ) pre λ=0 sa nazýva **derivácie funkcie $f(x)$ v bode x_o a v smere $\boldsymbol{n}$**

$$F'(0) = \operatorname{grad} f(\boldsymbol{x}_o) \cdot \boldsymbol{n} = \left| \operatorname{grad} f(\boldsymbol{x}_o) \right| \cos(\varphi)$$

kde φ je uhol, ktorý medzi smerom $\boldsymbol{n}$ a grad $f(\boldsymbol{x}_o)$



Geometrická interpretácia gradientu

Budeme študovať nasledujúce dva limitné prípady pre deriváciu v smere, a to (1) smer je paralelný s gradientom ($\varphi=0$) a (2) smer je antiparalelný s gradientom ($\varphi=\pi$)

$$F(0)\Big|_{\varphi=0} = +\left|\text{grad } f(\boldsymbol{x}_o)\right| > 0$$

$$F(0)\Big|_{\varphi=\pi} = -\left|\text{grad } f(\boldsymbol{x}_o)\right| < 0$$

V smere gradientu grad $f(\boldsymbol{x}_\text{o})$ funkcia $f(\boldsymbol{x})$ **najrýchlejšie rastie**, podobne, v opačnom smere - grad $f(\boldsymbol{x}_\text{o})$ funkcia $f(\boldsymbol{x})$ **najrýchlejšie klesá**.

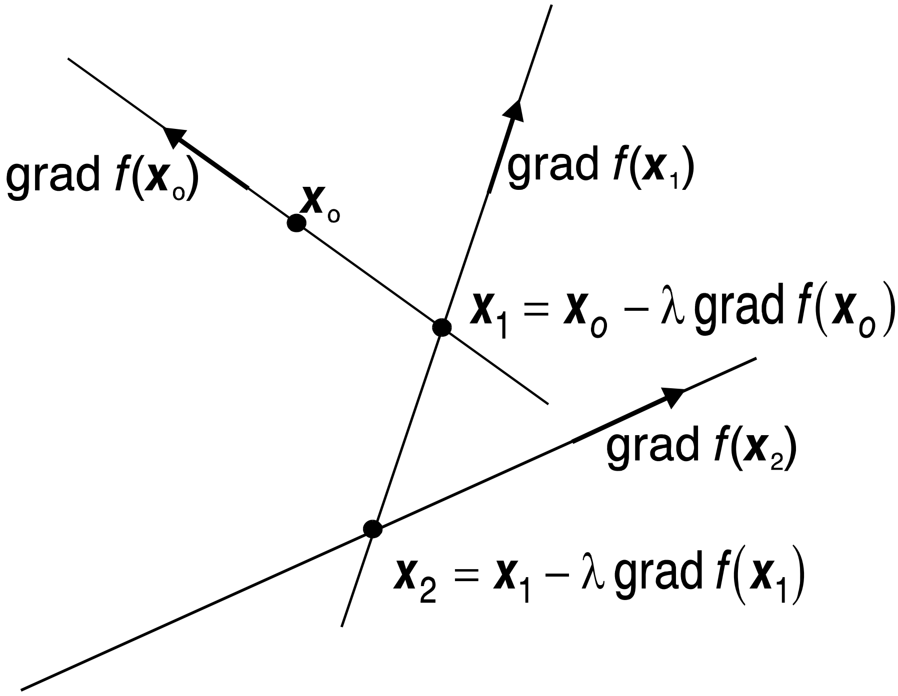
5. Metóda najprudšieho spádu (steepest desctent method)

Táto metóda je založená na geometrickej interpretácii gradientu, ako smeru v ktorom funkcia najrýchlejšie rastie. Budeme hľadať minimum funkcie $f(\boldsymbol{x})$, metóda najprudšieho spádu je založená na nasledujúcej rekurentnej formule

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \lambda \operatorname{grad} f(\boldsymbol{x}_k)$$

kde kladný parameter $\lambda > 0$ je určený tak, aby platilo $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$. To znamená, že tento parameter je určený dvoma protichodnými podmienkami, a to dostatočne malý, aby platila predchádzajúca nerovnosť a súčasne dostatočne veľký, aby bola zabezpečená dostatočná rýchlosť konvergencie metódy.

Grafická ilustrácia metódy najprudšieho spádu



Stratégia pre voľbu parametra λ

Ak platí $f(\mathbf{x}_{k+1})>f(\mathbf{x}_k)$ (t.j metóda prestáva monotónne konvergovať), potom parameter λ sa zmenší

$$\lambda \leftarrow \alpha \cdot \lambda$$

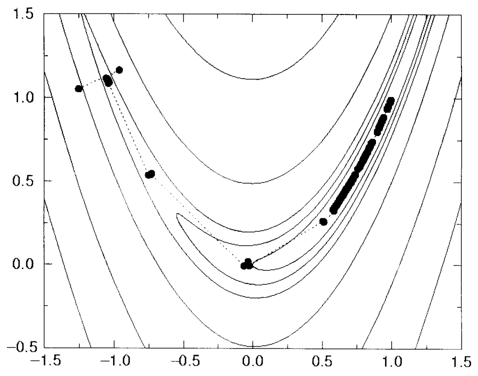
kde $0<\alpha<1$.

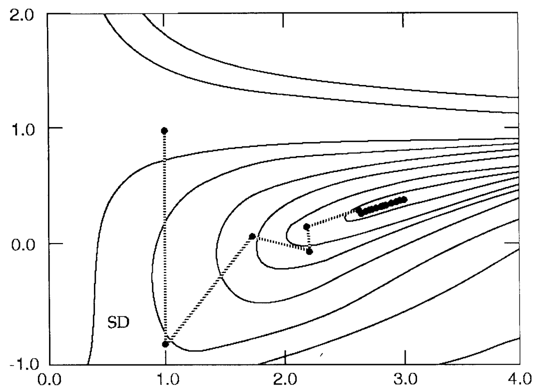
Algoritmus gradientovej metódy najprudšieho spádu

```
read(x,λ,ε,α,kmax);
norm:=∞; k:=0;
while (norm>ε) and (k<kmax) do
begin k:=k+1;
      x':=x-λ*gradf(x);
      norm:=|x-x'|;
      if f(x')>f(x) then
      λ:=α*λ;
      x:=x';
end;
write(x,f(x));
```

Grafická vizualizácia

metódy najprudšieho spádu





Metóda sa stáva veľmi pomalou v blízkosti minima alebo na "plochých" úsekoch.