



SPOLEHLIVOST KONSTRUKCÍ & TEORIE SPOLEHLIVOSTI

část 6: Zavedení statistické závislosti



Drahomír Novák
Jan Eliáš

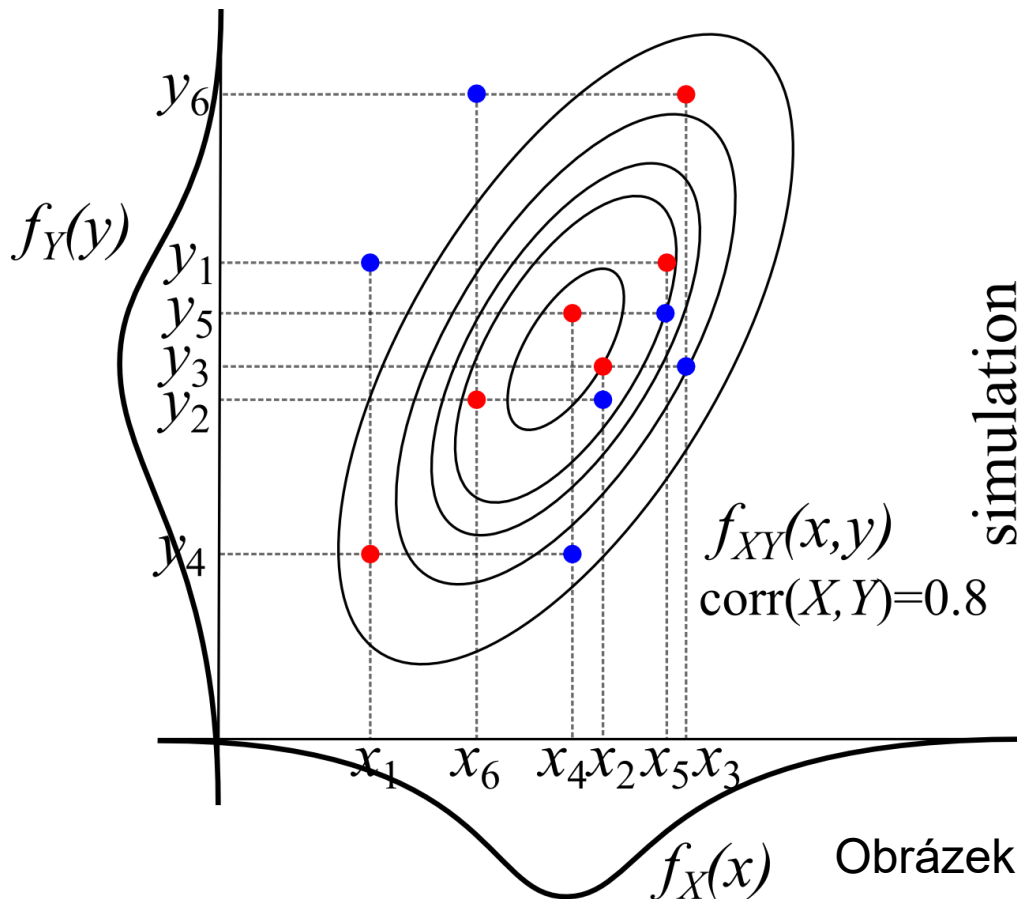


část 6

Zavedení statistické závislosti

Korelace mezi simulovanými veličinami

- zavede se pomocí **změny pořadí** jednotlivých realizací v simulační tabulce – shodné pro MC i LHS metodu



$\text{corr}(X, Y) = 0.0$

$\text{corr}(X, Y) = 0.8$

variable			variable		
simulation	1	x_1 y_1	simulation	1	$x_1 \rightarrow y_4$
	2	x_2 y_2		2	$x_2 \rightarrow y_3$
	3	x_3 y_3		3	$x_3 \rightarrow y_6$
	4	x_4 y_4		4	$x_4 \rightarrow y_5$
	5	x_5 y_5		5	$x_5 \rightarrow y_1$
	6	x_6 y_6		6	$x_6 \rightarrow y_2$
	\vdots	\vdots		\vdots	\vdots

Obrázek 1. Zavedení korelace v metodě MC



Korelační matice

- pro více než dvě náhodné vstupní veličiny je potřeba předepsat korelační koeficient pro **každou dvojici** veličin
- vhodná forma: korelační matice C
- vlastnosti korelační matice – **symetrická, pozitivě definitní**

	X_1	X_2	...	X_N
X_1	1	ρ_{12}	...	ρ_{1N}
X_2		1	...	ρ_{2N}
...
X_N	symmetric	1

- tento krok nelze přeskočit, **vždy je nějaká korelační matice použita**. Nuly mimo diagonálu jednoduše určují nulovou korelaci.

Obrázek 2. Korelační matice

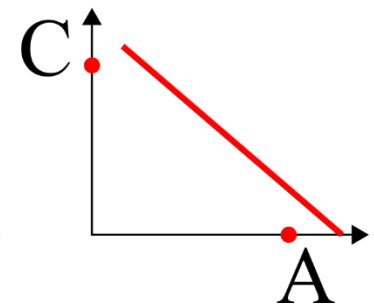
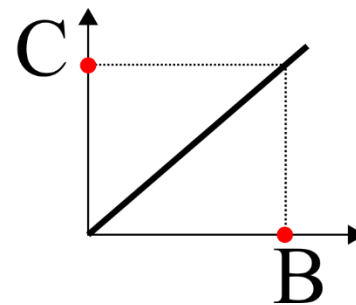
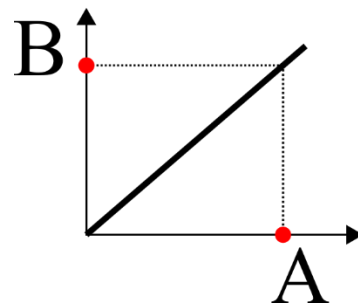
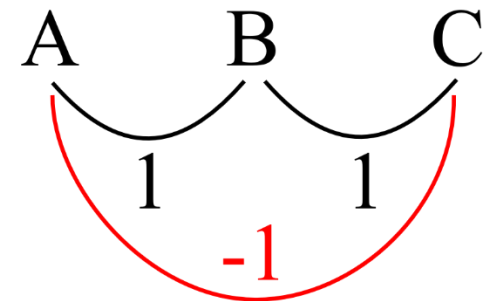


Pozitivní definitnost

- Znamená, že:
 - pro všechny x platí $x^T C x > 0$
 - všechna vlatní čísla jsou kladná
- v případě korelace to znamená, že matice je **smysluplná**

- příklad ukazuje **negativně definitní** korelační matici a z ní vyplývající **logický rozpor**

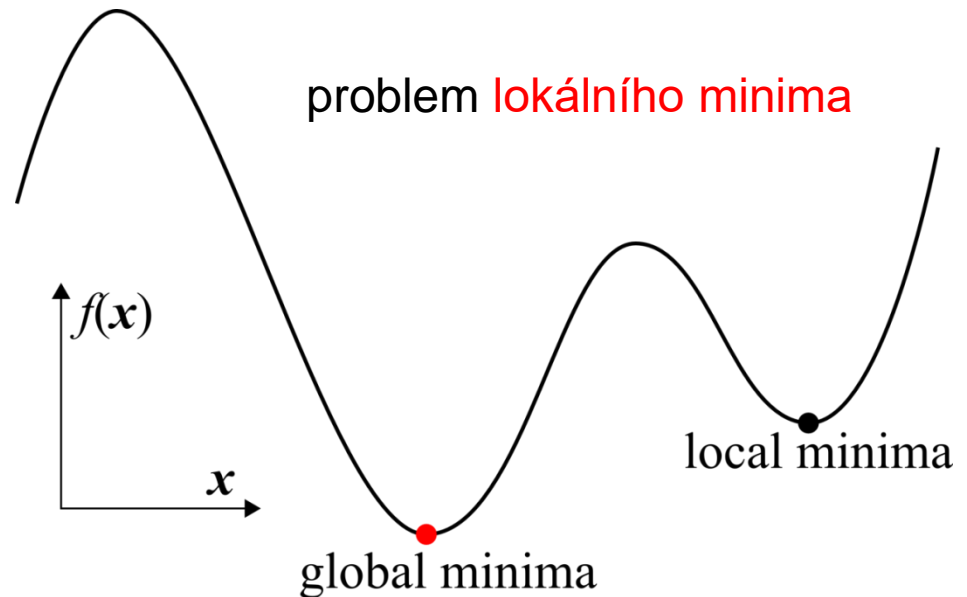
	A	B	C
A	1	1	-1
B	1	1	1
C	-1	1	1



Obrázek 3. Negativně-definitní korelační matice

Optimalizační problém

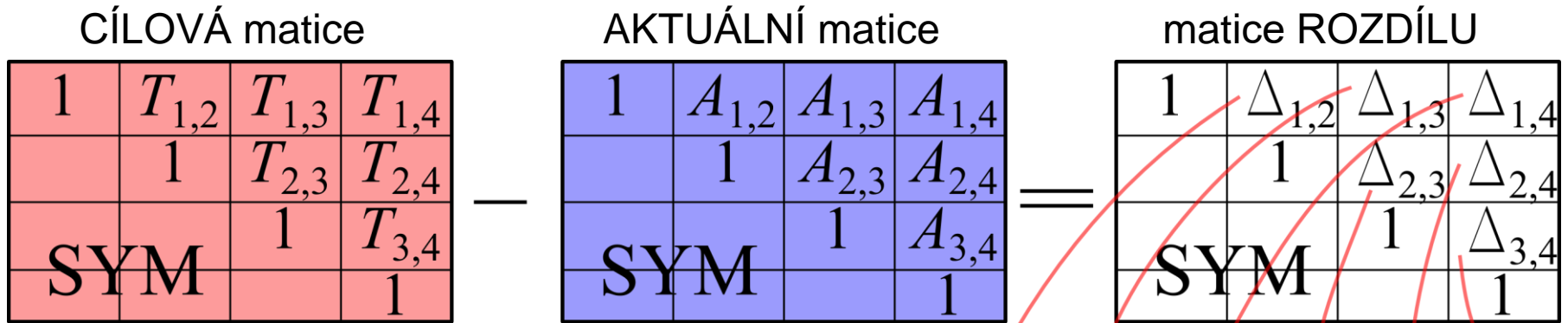
- mějme funkci f vektoru X na reálné číslo, chceme najít takový vektor x_o že pro všechna x platí: $f(x_o) \leq f(x)$
- funkce f se nazývá **cílová funkce**
- je možné požadovat další doplňující vlastnosti x_o souhrnně zvané **omezení**
- v našem případě hledáme přes všechny možná pořadí simulací to, které je nejbližší požadované **cílové korelační matici**



Obrázek 4. 1D objektivní funkce a její lokální a globální minimum

Cílová funkce

- rozdíl mezi cílovou a aktuální korelační maticí



Obrázek 5. matice rozdílu

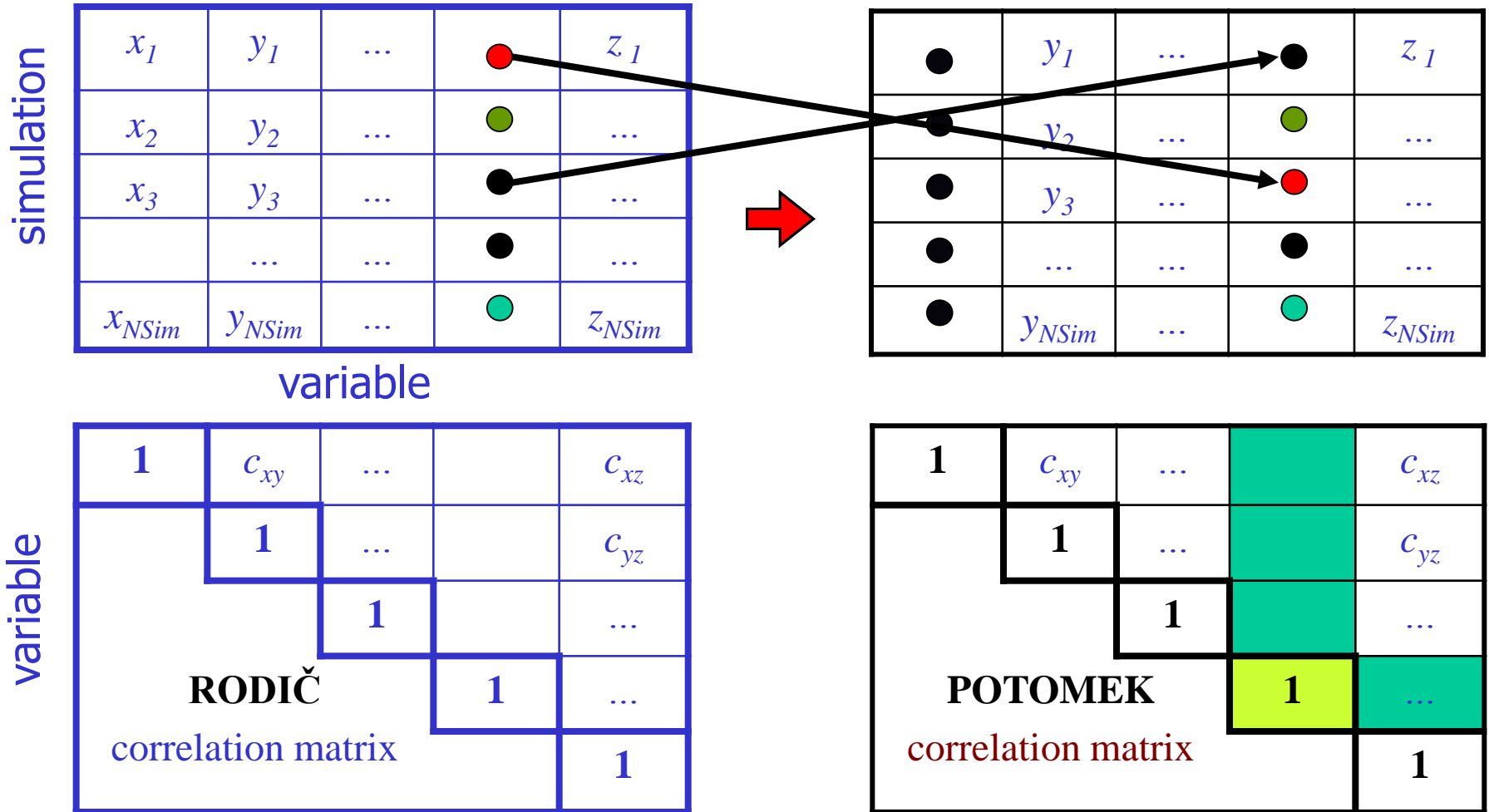
- **řádková norma** – maximální rozdíl v korelacích

$$f = Err = \max_{i=1}^{N_{cor}} |\Delta_i|$$

- **Euklidovská norma** – kvadratická

$$f = Err = \sqrt{\sum_{i=1}^{N_{cor}} \Delta_i^2}$$

Od “rodiče” k “potomkům”



Obrázek 6. Vytvoření potomka



Nejlepší pořadí

- ze všech možných kombinací pořadí
- je možné nalézt **globální minimum?**
- první sloupec zůstává, ostatní se míchají (permutují).
- existuje $(N_{Sim}!)^{N_{Var}-1}$ možných kombinací
- pro 6 simulací a 5 veličin je to celkem $(6!)^{5-1} = 2.6874 \cdot 10^{11}$ kombinací.

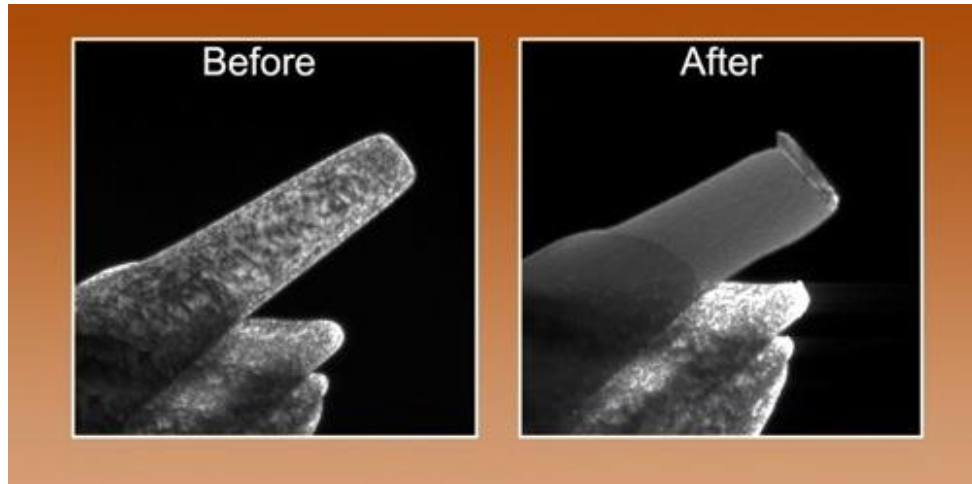
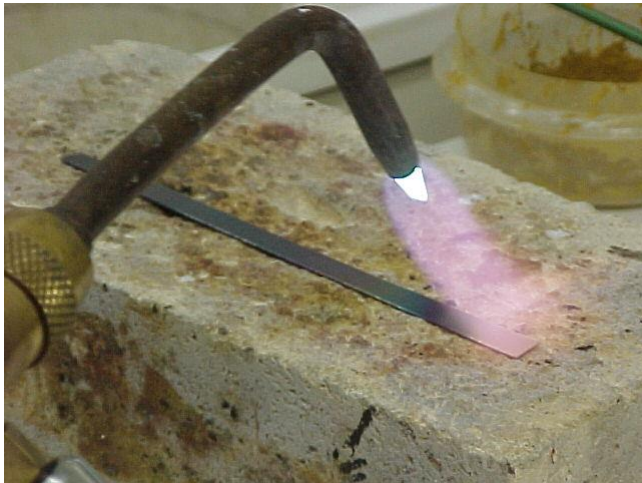
x_1	y_1	...	z_1
x_2	y_2	...	z_2
x_3	y_3	...	z_3
x_4	y_4	...	z_4
x_5	y_5	...	z_5
x_6	y_6	...	z_6
x_7	y_7	...	z_7
x_8	y_8	...	z_8
...
x_{NSim}	y_{NSim}	...	z_{NSim}

The diagram shows a table with 11 rows and 4 columns. The first column contains variables x_1 to x_{NSim} . The second column contains variables y_1 to y_{NSim} . The third column contains ellipses. The fourth column contains variables z_1 to z_{NSim} . A red arrow points from the top-right cell (z_1) to the bottom-right cell (z_{NSim}). A black arrow points from the second cell of the second row (y_2) to the second cell of the third row (y_3). A purple arrow points from the second cell of the fourth row (y_4) to the second cell of the fifth row (y_5). A green arrow points from the second cell of the sixth row (y_6) to the second cell of the seventh row (y_7). A black arrow points from the second cell of the eighth row (y_8) to the second cell of the ninth row (y_7).

Obrázek 7. Simulační tabulka

Optimalizační techniky

- hladový algoritmus (*akceptuje pouze zlepšení*)
- genetický algoritmus (*napodobuje evoluci*)
- optimalizace hejnem částic (*napodobuje sociální chování hejna*)
- optimalizace kolonií mravenců (*napodobuje chování mravenců*)
- **simulované žíhání** (*napodobuje žíhání krystalů*)



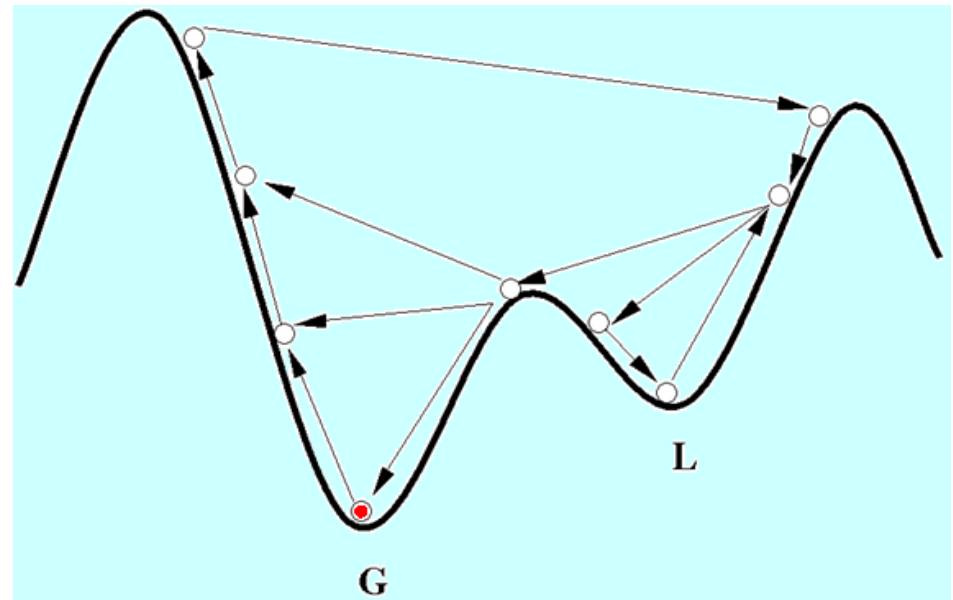
Obrázek 8. Žíhání kovů

Simulované žíhání

- stochastická optimalizační metoda
- vysoká pravděpodobnost nalezení **globalního minima**
- pravděpodobnost opuštění lokálního minima
- postupné chladnutí, pokles excitace systému

$$T_{i+1} = f_t T_i$$

- Boltzmann PDF,
analogie s energií
- vhodné i pro problémy
s dodatečnými omezeními
- konzistentní jednotky



Obrázek 9. Lokální a globální minimum

Algoritmus simulovaného žíhání

- pokud záměna pořadí **lepší** normu matice rozdílu ($\Delta E < 0$)

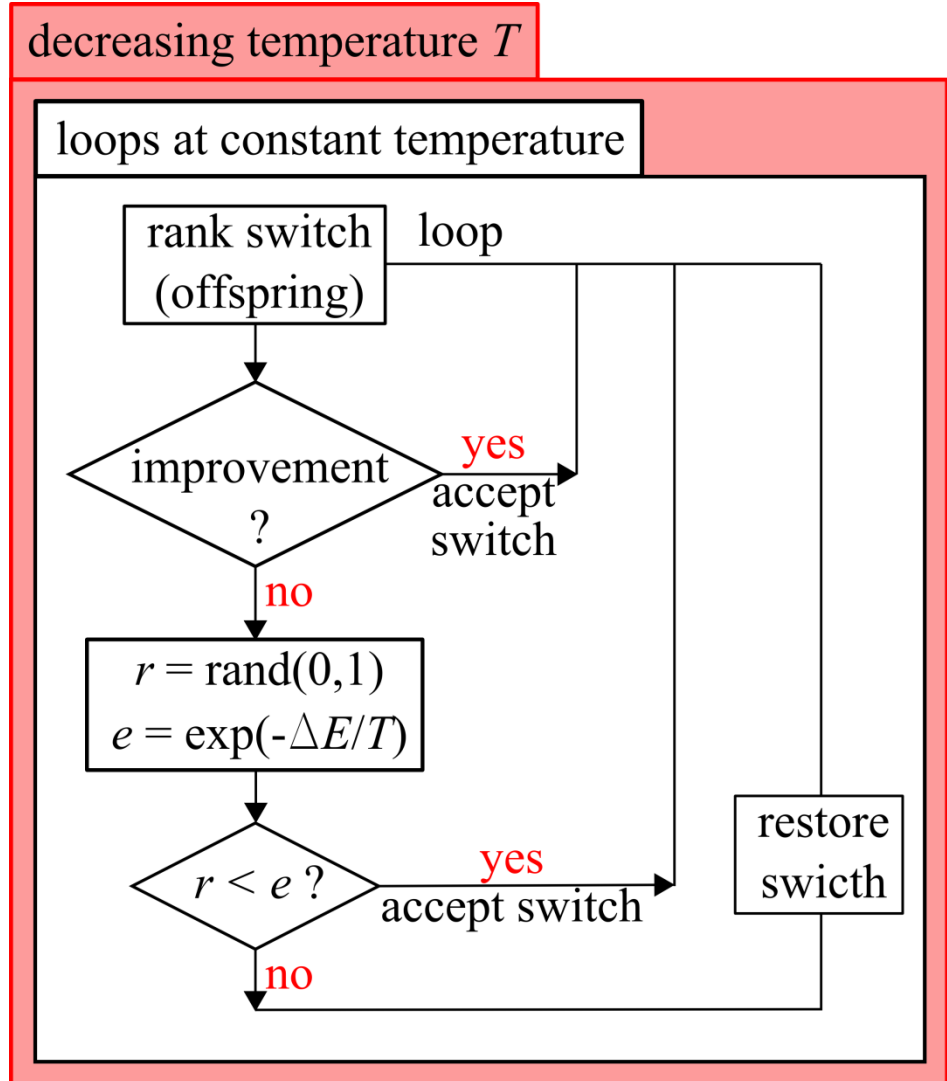
$$\Delta E = Err_i - Err_{i-1}$$

záměna je akceptována

- pokud ($\Delta E > 0$), **chyba roste**. Přesto, je akceptována s pravděpodobností

$$P_r(\Delta E) \approx e^{\left(\frac{-\Delta E}{k_b \cdot T}\right)}$$

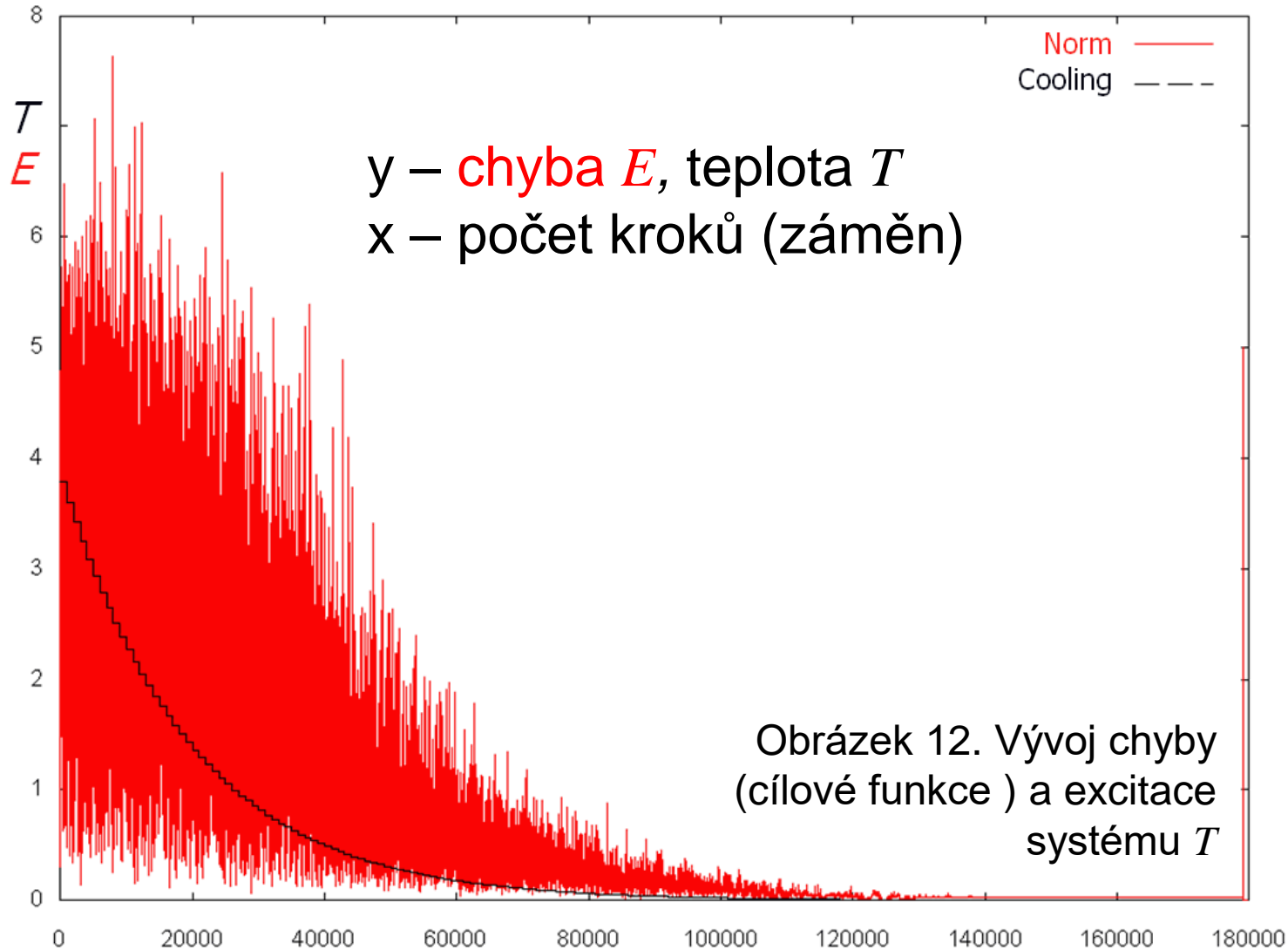
- to umožňuje uniknout z **lokálního minima**



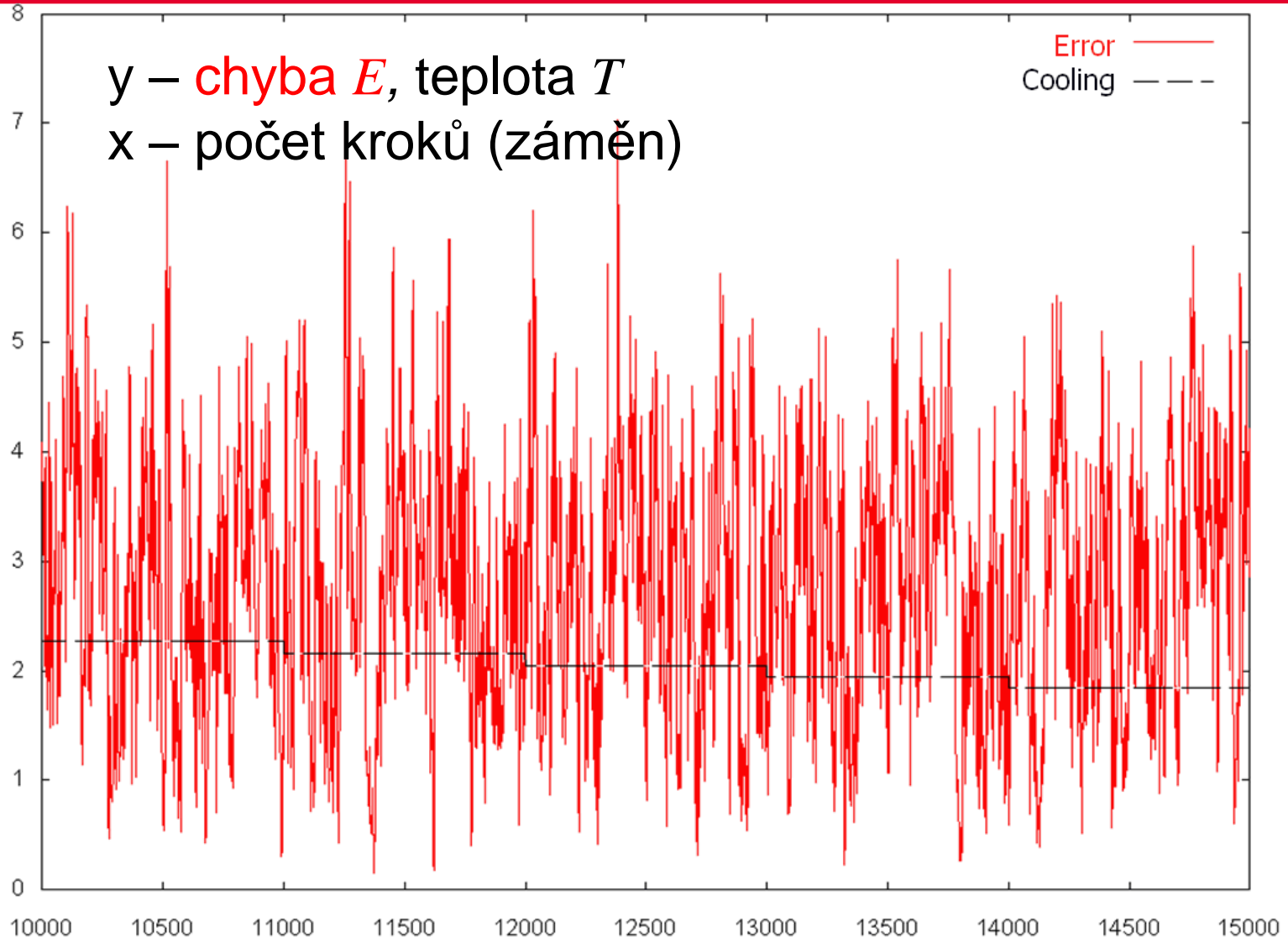
Obrázek 10. Algoritmus simulovaného žíhání

Vývoj chyby E

- průměrná chyba monotónně klesá
- rozptyl chyby klesá též (klesající excitace systému)



Vývoj chyby (detail)



Obrázek 12. Detail vývoje chyby



Test 1: Odstranění nežádoucí korelace

- 5 veličin ($N_{var}=5$), 6 simulací ($N_{sim}=6$)
- srovnáno s výsledky získanými Choleského dekomp., ULHS (Spearman)
- simulované žíhání, (PC 400MHz, 3 sec)

1				
0	1			
0	0	1		
0	0	0	1	
0	0	0	0	1

Obrázek 13. Cílová korelační matice testu 1



Test 1: Výsledky

Cholesky decomp. iterative ULHS (Spearman)

$$E_{overall}^2 = 0.22$$

1.4080	0.6867	0.2142	1.4080	-0.2142
-1.4080	-0.6867	1.4080	0.2142	0.2142
-0.6867	0.2142	-0.6867	0.6867	-1.4080
0.2142	-1.4080	-1.4080	-0.2142	0.6867
0.6867	-0.2142	0.6867	-1.4080	-0.6867
-0.2142	1.4080	-0.2142	-0.6867	1.4080

1	0.21	-0.20	0.07	-0.05
	1	0.11	0.12	0.11
		1	-0.08	-0.10
			1	-0.27
				1

target correlation matrix

1.4080
0.6867
0.2142
-0.2142
-0.6867
-1.4080

. < 0.1
. < 0.2
. < 0.3

Proposed algorithm (Simulated Annealing)

$$E_{overall}^2 = 0.04$$

0.2142	1.4080	0.6867	1.4080	0.2142
-1.4080	-0.6867	-0.2142	0.6867	-0.6867
0.6867	-1.4080	0.2142	0.2142	1.4080
-0.2142	0.2142	1.4080	-1.4080	-0.2142
1.4080	-0.2142	-0.6867	-0.2142	-1.4080
-0.6867	0.6867	-1.4080	-0.6867	0.6867

1	-0.10	0.06	-0.01	-0.08
	1	0.06	0.09	-0.09
		1	0	0.06
			1	0.05
				1

target correlation matrix

Obrázek 14. Výsledky testu 1

Test 2: Zavedení kladné korelace

- 5 veličin ($N_{var}=5$), 6 simulací ($N_{sim}=6$)
- simulované žíhání, (PC 400MHz, 3 sec)
- všechny možnosti = 50 min
- počet simulací má velký vliv na velikost chyby

1				
0.2	1			
0.2	0.6	1		
0.2	0.6	0.6	1	
0.2	0.5	0.2	0.5	1

Obrázek 15. Cílová korelační matice testu 2



Test 2: Výsledky

Starting sampling matrix

$$E_{overall}^2 = 3.79$$

1.4080	1.4080	1.4080	1.4080	1.4080
0.6867	0.6867	0.6867	0.6867	0.6867
0.2142	0.2142	0.2142	0.2142	0.2142
-0.2142	-0.2142	-0.2142	-0.2142	-0.2142
-0.6867	-0.6867	-0.6867	-0.6867	-0.6867
-1.4080	-1.4080	-1.4080	-1.4080	-1.4080

1	1.0	1.0	1.0	1.0
0.2	1	1.0	1.0	1.0
0.2	0.6	1	1.0	1.0
0.2	0.6	0.6	1	1.0
0.2	0.5	0.2	0.5	1

1.4080
0.6867
0.2142
-0.2142
-0.6867
-1.4080

. < 0.1
. < 0.2
. < 0.3

Proposed genetic algorithm (Simulated Annealing)

$$E_{overall}^2 = 0.004$$

1.4080	1.4080	1.4080	1.4080	1.4080
0.2142	-0.2142	-0.2142	-1.4080	-0.6867
-0.6867	-0.6867	0.6867	0.2142	-1.4080
-1.4080	0.2142	-0.6867	-1.4080	-0.2142
1.4080	0.6867	0.2142	0.6867	0.6867
0.6867	-1.4080	-1.4080	-0.6867	0.2142

1	0.165	0.237	0.203	0.517
0.2	1	0.590	0.588	0.537
0.2	0.6	1	0.615	0.193
0.2	0.6	0.6	1	0.503
0.2	0.5	0.2	0.5	1

Obrázek 16. Výsledky testu 2

Vážená optimalizace

1	0,9	0,9
	1	-0,9
		1

→
optim

1	0,49	0,49
	1	-0,49
		1

–výsledná korelační matice je pozitivně definitní a chyba je rovnoměrně rozdělena mezi všechny korelační koeficienty

1	0,9	0,9
	1	-0,9
		1

→
optim

1	0,2	0,2
1	1	-0,9
1	100	1

–**vážená metoda**: potlačení vybraných koeficientů



Citlivostní analýza

- dokáže určit **dominující** a **nedominující** vstupní náhodné veličiny
- neparametrická pořadová korelace mezi **vstupními** veličinami X_i a **výstupy** G
- dodatečný výsledek LHS (či MC) metody, není nutné provádět žádné další analýzy
- relativní vliv je vyjádřen korelačními koeficienty v rozmezí **(-1, 1)** - Kendallovo tau či Spearmanův korelační koeficient
- robustní – použití pouze pořadí
- **“větší” korelační koeficient = vyšší citlivost**



Citlivostní analýza

– neparametrická pořadová korelace mezi vstupními a výstupními veličinami

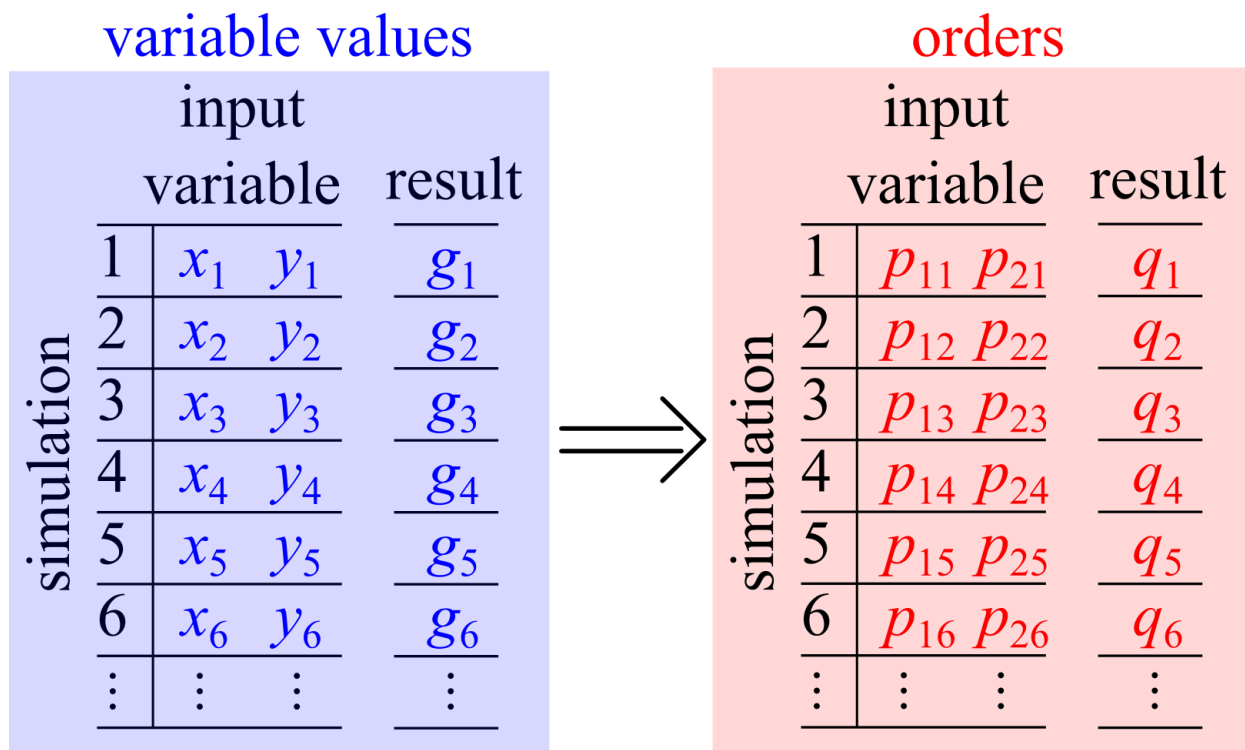
- Spearmanův korelační koeficient

$$\rho_i^s = 1 - \frac{6 \sum_{j=1}^{N_{sim}} (p_{ij} - q_j)^2}{N_{sim} (N_{sim}^2 - 1)}$$

- Kendall tau

$$\tau_i = \tau(p_{ij}, q_j),$$

$$j = 1, 2, \dots, N_{sim}$$



Obrázek 18. Simulované hodnoty a pořadí veličin