Lévy flights in binary optimization

MARTIN KLIMT, JAROMíR KUKAL and MATEJ MOJZEŠ

There are many optimization heuristics which involves mutation operator. Reducing them to binary optimization allows to study properties of binary mutation operator. Modern heuristics yield from Lévy flights behavior, which is a bridge between local search and random shooting in binary space. The paper is oriented to statistical analysis of binary mutation with Lévy flight inside and Quantum Tunneling heuristics.

Key words: binary optimization, mutation operator, Lévy flights, Bernoulli trials, Quantum Tunneling

1. Binary optimization

Let $n \in \mathbf{N}$ be dimension, $\mathbf{D} = \{0,1\}^n$ be domain of objective function $\varphi : \mathbf{D} \to \mathbf{R}$. *Binary optimization* task is defined as finding of any $\mathbf{x}_{opt} \in \mathbf{D}$ satisfying

$$\mathbf{x}_{opt} \in \arg\min_{\mathbf{x} \in \mathbf{D}} \varphi(\mathbf{x}). \tag{1}$$

There are many heuristics for solving the task (1). Genetic Optimization (GO) [1], Simulated Annealing (SA) [2], Fast Simulated Annealing (FSA) [3], Quantum Tunneling (QT) [4], Discrete Cuckoo Search (DCS) [5], and Discrete Firefly Search (DFS) [6] are frequently used in binary optimization. Basic operator of binary optimization is mutation which avoids trap in local minima (SA, FSA, QT) or population degeneration (GO, DCS, DFS).

2. Binary mutation

Binary mutation can be defined by stochastic function $\mu : \mathbf{D} \to \mathbf{D}$ satisfying $\mu(\mathbf{x}) \neq \mathbf{x}$ for all $\in \mathbf{D}$. New vector is generated by mutation as

$$\mathbf{x}_{new} = \mu(\mathbf{x}) \tag{2}$$

The Author is with Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague Department of Software Engineering, Trojanova 13, Praha 2, 120 00 Czech Republic. Corresponding author is J. Kukal, e-mail: jaromir.kukal@fjfi.cvut.cz

which can be decomposed to

$$\mathbf{x}_{new} = (\mathbf{x} + \mathbf{y}) \bmod 2 \tag{3}$$

where

$$\mathbf{y} = \mu(\mathbf{0}) \neq \mathbf{0}. \tag{4}$$

Therefore, any binary mutation can be studied as mutation from the origin $\mathbf{0}$. Any reasonable binary mutation is invariant to $\mathbf{x}$, $\mathbf{x}_{new}$ coordinate permutation. This is main reason for why we use stochastic variable

$$k = \|\mathbf{x}_{new} - \mathbf{x}\|_{H} = \|\mathbf{y}\|_{H} \in \{1, ..., n\} \tag{5}$$

for representation of any binary mutation. Here, $\|...\|_{H}$ means Hamming norm and knowledge of k value enables generation of $\mathbf{y}$ as one of $\binom{n}{k}$ vectors with just k unit coordinates. Using Minkowski norm with $p \geqslant 1$, we easily recognize that

$$\|\mathbf{y}\|_{p} = \|\mathbf{y}\|_{H}^{1/p} \tag{6}$$

which is monotonically increasing function of parameter p and then we does not obtain alternative view to binary mutation description. Therefore, any binary mutation can be characterized by probabilities

$$\{p_k\}_{k=1}^{n}, \quad p_k \geqslant 0, \quad \sum_{k=1}^{n} p_k = 1 \tag{7}$$

which describes stochastic variable k in (5) and then any binary mutation.

Traditional Mutation (TRM) is based on Bernoulli trials [7] with event probability $p_{mut} \in (0, ^1\!/_2)$, event number n, and excluded value $k = 0$. Resulting probabilities are

$$p_k = \binom{n}{k} \cdot \frac{p_{mut}^{k}(1 - p_{mut})^{n-k}}{1 - (1 - p_{mut})^{n}} \tag{8}$$

When $p_{mut} \to ^1\!/_2-$, traditional mutation approaches *Random Shooting Mutation* (RSM) to another point with probabilities

$$\lim_{p_{mut} \to ^1\!/_2-} p_k = \binom{n}{k} \cdot (2^n - 1)^{-1}. \tag{9}$$

Second extreme, *Nearest Neighbor Mutation* (NNM) occurs when $p_{mut} \to 0+$, then

$$\lim_{p_{mut} \to 0+} p_k = \delta_{k,1} \tag{10}$$

and stochastic variable degenerates to $k = 1$.

If TRM has $p_{mut} \leqslant 2/(n+1)$, series $\{p_k\}_{k=1}^{n}$ is non-increasing one, having only one maximum, and one minimum. When $p_{mut} > 2/(n+1)$, the series $\{p_k\}_{k=1}^{n}$ has two local minima plus single maximum for $k > 1$. That is why TRM is only successful for small p_{mut} which is just useful for local searching.

3. Continuous random variable as generator of binary mutation

Let $\xi \in \mathrm{R}^n$ be continuous random variable with density $f : \mathrm{R}^n \to \mathrm{R}_0^+$ satisfying $\int f(\xi) \, d\xi = 1$ and $f(-\xi) = f(\xi)$ for all $\xi \in \mathrm{R}^n$. Let $T > 0$ be parameter which is called *temperature*. Binary mutation (4) can be generated by repeating

$$\mathbf{y} = \lfloor T\xi + {}^1\!/_2 \rfloor \mod 2 \tag{11}$$

until $\mathbf{y} \neq \mathbf{0}$. This special approach can bring new mutation class with better properties than TRM.

4. Negative result for independent coordinates

Supposing independent coordinates of vector ξ we can decompose its density as

$$f(\xi) = \prod_{k=1}^{n} g(\xi_k) \tag{12}$$

and then generate individual binary coordinates as

$$y_k = \lfloor T\xi_k + {}^1\!/_2 \rfloor \mod 2. \tag{13}$$

This independent approach brings unpleased result for any PDF g and its CDF G because

$$prob(y_k = 1) = \sum_{m \in \mathrm{Z}} \left(G\left(\frac{2m + {}^3\!/_2}{T}\right) - G\left(\frac{2m + {}^1\!/_2}{T}\right) \right) \tag{14}$$

does not depend on index k and then (13) only generates Bernoulli trials with

$$p_{mut} = prob(y_k = 1) = q(T) \tag{15}$$

where $q(T)$ is non-decreasing function of temperature T for any PDF g and its CDF G according to (14). Therefore, distributions with independent coordinates are not useful for realization of binary mutation which is different from TRM.

5. Lévy flight as source of mutation

Let $\beta \in (0, 2)$ be parameter of Lévy distribution [8], let $d \sim L(\beta)$ be stochastic variable with Lévy distribution, and let $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$, $\mathbf{z} \in \mathrm{R}^n$ be stochastic variable with multivariate Gaussian distribution.

Lévy Flight from origin is then defined as generation of stochastic variable

$$\xi = \frac{\mathbf{z}}{\|\mathbf{z}\|_2} \cdot d. \tag{16}$$

Adequate PDF $f(\xi)$ is radial symmetric but coordinates of ξ are fortunately dependent and the PDF has heavy tails. Random variable with Lévy distribution can be generated by Mantegna formula [9]

$$d = \frac{u}{|v|^{1/\beta}} \tag{17}$$

where $u \sim N(0, \sigma^2)$, $v \sim N(0, 1)$ and

$$\sigma = \left(\frac{\Gamma(1+\beta)\sin(\pi\beta/2)}{\Gamma((1+\beta)/2) \cdot \beta \cdot 2^{(\beta-1)/2}} \right)^{1/\beta}. \tag{18}$$

After binarization (11), we can generate *Lévy Flight Mutation* (LFM), study its stochastic variable k, and influence of temperature T for fixed n, β.

When T is small, p_k of LFM monotonically decreases and then local search is preferred as in TRM. When T is large, p_k of LFM is also similar to TRM and near to RSM. But for medium temperature T, p_k of LFM has two maxima, which is a kind of hybridization between local and global search. LFM behavior was studied using Monte Carlo technique [10] for $n = 20$, $\beta = {}^3\!/_2$ 10^6 trials, and various temperatures. Resulting probabilities are depicted in Figs. 1-3.

6.   Mutation with parasitism

Let $w \in (0, 1)$ be probability of parasitism, $\mathbf{y}_{TRM}$ be result of traditional mutation, and $\mathbf{y}_{RSM}$ be result of random shooting. *Mutation with parasitism* (PAM) is defined here as

$$\mathbf{y} = \begin{cases} \mathbf{y}_{RSM} & \text{with probability} \quad w \\ \mathbf{y}_{TRM} & \text{with probability} \quad 1-w. \end{cases} \tag{19}$$

Despite of PAM triviality, this type of binary mutation has similar properties as LFM. PAM behavior was studied using Monte Carlo technique for $n = 20$, $p_{mut} = 0.05$, 10^6 trials, and various parasitism probabilities and the results were depicted in Figs. 4-6. The main question is not about similarity of LFM and PAM distributions, but about LFM and PAM applicability in the case of difficult binary optimization tasks.

7.   Experimental study: Quantum Tunneling for prime number problem

Quantum Tunneling (QT) [4] is physically motivated heuristics based on LFM. Beginning with $k = 0$, $T_k > 0$, and

$$\mathbf{x}_0 \sim U(\mathbf{D}) \tag{20}$$

we perform LFM with temperature T_k to obtain

$$\mathbf{y}_k = LFM(\mathbf{x}_k, T_k, \beta) \tag{21}$$

and move to the better state according to rule

$$\mathbf{x}_{k+1} = \begin{cases} \mathbf{y}_k, & f(\mathbf{y}_k) < f(\mathbf{x}_k) \\ \mathbf{x}_k, & f(\mathbf{y}_k) \geqslant f(\mathbf{x}_k) \end{cases} \tag{22}$$

until termination condition holds. The cooling strategy is represented by non-increasing sequence of positive temperatures T_k. We used constant temperature $T_k = T_0$ and termination condition

$$f(\mathbf{x}_k) \leqslant f^* \ \text{ or } \ k \geqslant maxeval. \tag{23}$$

Respectively, we can use PAM instead of LFM in (21) as

$$\mathbf{y}_k = PAM(\mathbf{x}_k, p_{mut}, w) \tag{24}$$

and compare time complexity of QT with LFM or alternative PAM inside.

There are many possibilities how to design difficult but reasonable binary optimization task with many local extremes. *Prime Number Problem* (PNP) can be defined in this paper as finding maximum prime number not exceeding N in binary representation of course, which can be formulated as constrained optimization task

$$q = \max \tag{25}$$

$$q = \sum_{k=1}^{n} 2^{k-1} x_k \tag{26}$$

$$\mathbf{x} \in \{0,1\}^n \tag{27}$$

$$q \in \mathrm{P} \tag{28}$$

$$q \leqslant N \tag{29}$$

where $\mathbf{P}$ is set of prime numbers. Using penalty technique, we transformed PNP to n-dimensional binary optimization task with multimodal objective function

$$\Phi(\mathbf{x}) = N - q + 2^n \cdot (\max(0, q-N) + 1 - (q \in \mathbf{P})) \tag{30}$$

which was subject of minimization. When $N = 10^6$, $n = 20$, then global minimum has the value $\Phi(\mathbf{x}^*) = 17$. It corresponds to prime number 999983 and will be used for optimization experiments.

Numerical testing of QT with LFM/PAM inside on PNP task was performed with various mutation parameters. Every parameter setting was investigated via 100 independent runs with maximum number of evaluations per run set to 100000. According to tradition, we evaluate *ene* as mean number evaluations in the case of successful optimization, *rel* as reliability of optimum searching [%], and $FEO = ene/rel$ as Feoktistov criterion [11] as widely recommended complexity measure. Results of computer experiments for QT with LFM inside are collected in Tab. 1. Time complexity FEO is relatively independent on parameter β. Optimum temperature is $T_0 = 2$ in this case. Results

Table 11. Time complexity of QT with LFM in the case of PNP

β	0.5			1			1.5		
T_0	*ene*	*rel*	*FEO*	*ene*	*rel*	*FEO*	*ene*	*rel*	*FEO*
0.5	54644	43	127080	46843	43	108938	55871	29	192659
1	46842	43	108935	47343	49	96619	40160	33	121697
2	51342	44	116687	51835	54	95991	52866	54	97900
5	52794	36	146650	52249	45	116109	55156	48	114913
10	48402	40	121005	54887	36	152464	46128	35	131790

Table 12. Time complexity of QT with PAM in the case of PNP

β	0.05			0.1			0.2		
T_0	*ene*	*rel*	*FEO*	*ene*	*rel*	*FEO*	*ene*	*rel*	*FEO*
0.3	46069	23	200300	48120	30	160400	48848	45	108552
0.4	49792	27	184415	41265	27	152834	49252	58	84918
0.5	56621	24	235921	50742	29	174973	50795	56	90642
0.8	50865	20	254325	44198	38	116311	60187	60	100312
0.9	49578	29	170959	52311	25	209244	60102	37	162438

of computer experiments for QT with PAM inside are collected in Tab. 2, where lower values of mutation probability have to be compensated by higher values of parasitism. Time complexity FEO has optimum for $p_{mut} = 0.2$, $w = 0.5$. As seen, parasite mutation achieved lower time complexity criterion FEO than mutation with Lévy flights in this case, which motivates us to substitute LFM by PAM in binary optimization heuristics.

8. Conclusions

Binary mutation operator as investigated both theoretically and numerically. Traditional mutation with Bernoulli trials is ineffective in general, but binary mutation inspired by Lévy flights helps to realize trade off between local and global searches. Similar properties were detected also in the case of binary mutation with parasitism, which is a statistical mixture of Bernoulli trials and random shooting. As demonstrated on twenty-dimensional Prime Number Problem, Quantum Tunneling heuristics offered relative low
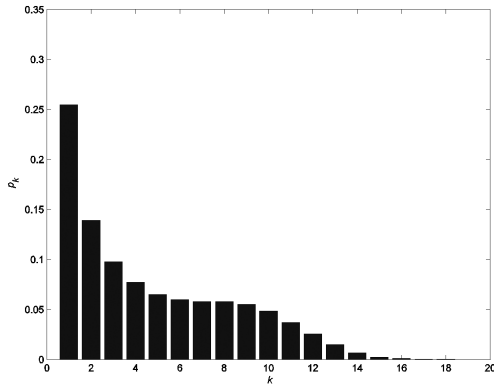
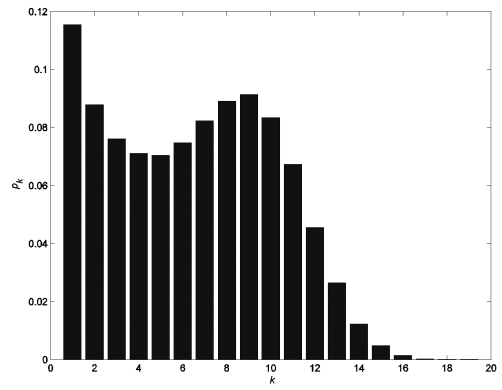Figure 1. LFM for $T = 0.5$.

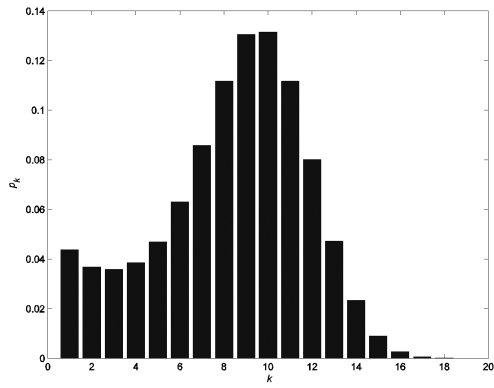Figure 2. LFM for $T = 2$.

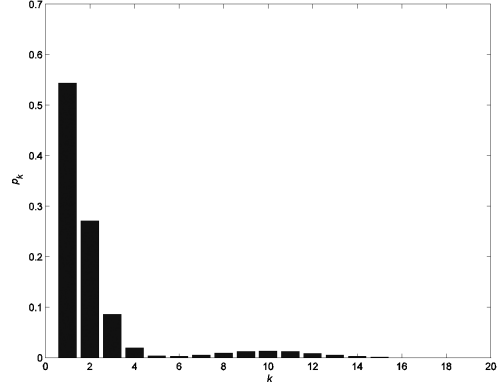Figure 3. LFM for $T = 5$.

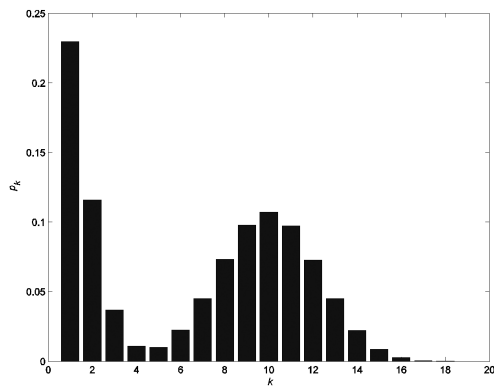Figure 4. PAM for $w = 0.05$.
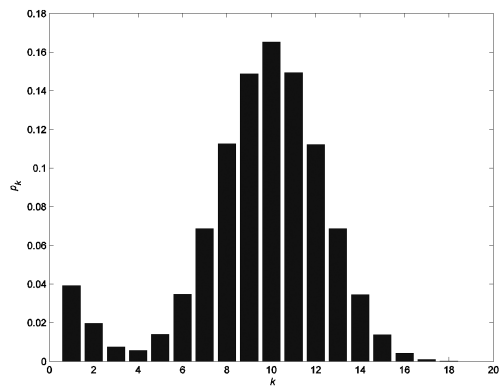

Figure 5. PAM for $w = 0.5$.

Figure 6. PAM for $w = 0.9$.

time complexity value for both Lévy flights and parasitic mutation, which is a very good inspiration for the other tasks and heuristics of binary optimization.

References

[1] D.E. GOLDBERG: Genetics Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.

[2] V. ČERNÝ: Thermodynamical approach to the travelling salesman problem: An efficient simulation algorithm. *J. of Optimization Theory and Applications*, **45**(1), (1985), 41-51.

[3] H. SZU and R. HARTLEY: Fast simulated annealing. *Physics Letters A*, **122**(3-4), (1987), 157-162.

[4] B. APOLLONI, C. CARAVALHO and D. DE FALCO: Quantum stochastic optimization. *Stochastic Processes and their Applications*, **33**(1), (1989), 233-244.

[5] X. OUYANG, Y. ZHOU, Q. LUO and H. CHEN: A novel discrete cuckoo search algorithm for spherical traveling salesman problem. *Applied mathematics & information sciences*, **7**(2), (2013), 777-784.

[6] M.K. SAYADI, R. RAMEZANIAN and N. GHAFFARI-NASAB: A discrete firefly meta-heuristic with local search for makespan minimization in permutation flow shop scheduling problems. *Int. J. of Industrial Engineering Computations*, **1**(1), (2010), 1-10.

[7] R. MOTWANI and P. RAGHAVAN: Randomized Algorithms. Cambridge University Press, 1995.

[8] G.M. VISWANATHAN: Lévy flights and superdiffusion in the context of biological encounters and random searchers.*Physics of Life Reviews*, **5**(1), (2008), 133-150.

[9] R.N. MANTEGNA: Fast, accurate algorithm for numerical simulation of Lévy stable stochastic processes. *Physical Review E*, **49**(5), (1994), 4677-4683.

[10] D.P. KROESE, T. TAIMRE and Z.I. BOTEV: Handbook of Monte Carlo Methods. John Wiley & Sons, 2011.

[11] V. FEOKTISTOV: Differential Evolution in Search of Solutions, Springer, 2006.