Integer Heuristic Optimization

Matej Mojzeš

Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University in Prague
Trojanova 13, 120 00 Prague, Czech Republic

December 7 2017

Outline

1. Time complexity measures – from extended search process to quantile and other measures
2. Novel heuristics – how to apply Lévy Flights and competitive approach in integer optimization
3. Selected applications – general linear models and submodel selection via binary optimization
4. Experimental results – selection and tuning of heuristics on selected tasks, incl. submodel optimization in fractography and biomedicine

Part I: Time Complexity Measures

- Integer Optimization Task
- Extended Search Process
- Quantile Time Complexity Measure
- Mean Value Time Complexity Measures
- Applicability
- Details in:
  M. Mojzeš and J. Kukal. Quantile and Mean Value Measures of Search Process Complexity. *Journal of Combinatorial Optimization*, In press

Integer Optimization Task (IOT)

- Domain of optimization

$$\mathcal{D} = \{\mathbf{x} \in \mathbb{Z}^d : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\} \tag{1}$$

- Objective function f: $\mathcal{D} \rightarrow \mathbb{R}$
- Target value selection

$$f^* \geq \min_{\mathbf{x} \in \mathcal{D}} \mathrm{f}(\mathbf{x}) \tag{2}$$

- Goal set specification

$$\mathcal{G} = \{\mathbf{x} \in \mathcal{D} : \mathrm{f}(\mathbf{x}) \leq f^*\} \tag{3}$$

Search Process (SP)

- Maximum number of searching steps: $N \in \mathbb{N}$
- SP: any algorithm generating $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) \in \mathcal{U}$
- In the case of IOT:
  - $\mathcal{U} = \mathcal{D}$
  - $n = \min\{n : \mathbf{x}_n \in \mathcal{G}\}$
- Integer Optimization Heuristic (IOH)
- SP = IOT + IOH in this case

Extended Search Process (XSP)

- Should the SP end with a failure, we proceed with new runs until succeeding
- This trivial but practical habit can be called XSP
- Example of an XSP:

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_1'$	$\mathbf{x}_2'$	$\mathbf{x}_3'$	$\mathbf{x}_4'$	$\mathbf{x}_1''$	$\mathbf{x}_2''$	$\mathbf{x}_3''$

- $N = 4,\ H = 2,\ j = 3,\ \mathrm{f}(\mathbf{x}_3'') \leq f^*,\ \mathbf{x}_3'' \in \mathcal{G}$
- $1 \leq j \leq N,\ n^* = NH + j$

Novel Quantile Measure

- For $0 < Q < 1$ we define

$$n_Q^* = \min\{n : F_n \geq Q\} \tag{4}$$

- which can be calculated as

$$n_Q^* = NH + j \tag{5}$$

- where the number of full runs is

$$H = \left\lceil \frac{\ln\left(\frac{1-Q}{1-F_N}\right)}{\ln(1-F_N)} \right\rceil \tag{6}$$

- with the additional evaluation number

$$j = \left\lceil \mathrm{F}^{-1}\left(1 - \frac{1-Q}{(1-F_N)^H}\right) \right\rceil \leq N \tag{7}$$

Mean Value Time Complexity Measures

- Mean complexity measure is defined as

$$\mathrm{E}\,\Phi(n^*) = \int_0^1 \Phi(n_Q^*) \cdot \mathrm{d}Q \tag{8}$$

- and approximated as

$$\mathrm{E}\,\Phi(n^*) \approx \frac{1}{M} \cdot \sum_{k=0}^{M-1} \Phi\left(n^*_{\frac{k+1/2}{M}}\right) \tag{9}$$

- Particular cases:
  1. Mean time complexity $CRIT_1 = \mathrm{E}\,n^*$ in formal agreement with (Hoos and Stutzle, 1998) despite using different statistical reasoning

$$CRIT_1 = SP2 = \left(\frac{1-p_s}{p_s}\right) FE_{\max} + \mathrm{E}\,T_A^s \tag{10}$$

  2. Mean logarithmic complexity $CRIT_2 = \mathrm{E} \ln n^*$
  3. Mean rate of search $CRIT_3 = \mathrm{E}(n^*)^{-1}$

Applicability to a given IOT only

- No Free Lunch Theorem for Optimization holds
- Complexity measures will be applied to a separate IOT
  - When the IOT solution is unknown, we can still tune IOH on similar/smaller IOTs.
- Given IOH
  - Given parameters $\rightarrow N_{\text{opt}}$
  - Given $N \rightarrow$ optimal parameters
- Comparison of various IOH using their best settings for a given N

Part II: Rank Mean Field Integer Flight (RMFIF)

- Novel IOH
- Based on the mean field theory
- Rank of objective function values
- Lévy Flight mutation
- Competitive approach
- Optional hybridization
- Details in:
  M. Mojzeš and J. Kukal. Competitive Rank Based Discrete Optimization Metaheuristic with Lévy Flights. *Soft Computing*, Under review.

Mean Field Theory (MFT) on Sorted Population

- Sorted population

$$\mathbf{P} = ((\mathbf{x}_1, f_1), \ldots, (\mathbf{x}_N, f_N)) \ . \tag{11}$$

- MFT is based on the partition function

$$Z = \sum_{k=1}^{N} \exp(-k/T_{\mathrm{MFT}}) \tag{12}$$

- Steady state probabilities are

$$p_k = \exp(-k/T_{\mathrm{MFT}})/Z = \frac{1-Q}{1-Q^N} \cdot Q^{k-1} \tag{13}$$

where $Q = \exp(-T_{\mathrm{MFT}}^{-1})$

Anisotropic Mutation with Lévy Flight

- Population center via MFT

$$\mathbf{e} = \mathrm{E}\mathbf{x} = \sum_{k=1}^{N} p_k \mathbf{x}_k \tag{14}$$

- Covariance matrix via MFT

$$\mathbf{C}_{\mathrm{raw}} = \sum_{k=1}^{N} p_k (\mathbf{x}_k - \mathbf{e})(\mathbf{x}_k - \mathbf{e})^{\mathrm{T}} \tag{15}$$

- Anisotropic directional vector

$$\mathbf{z} = (\mathbf{C}_{\mathrm{raw}} + \sigma^2 \mathbf{I})^{1/2} \cdot (\mathbf{y}/\|\mathbf{y}\|_2) \tag{16}$$

where $\mathbf{y} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

Anisotropic Mutation with Lévy Flight – continuation

- Real unconstrained mutation

$$\mathbf{y} = \mathbf{x} + T_{\mathrm{mut}} \cdot \mathbf{z} \cdot \tan \frac{\pi r}{2} \tag{17}$$

  where $r \sim \mathrm{U}(0, 1)$
- Integer unconstrained mutation $\mathbf{z} = \lfloor \mathbf{y} + 1/2 \rfloor$
- Necessary perturbation $\mathbf{x}_{\mathrm{new}} = \mathrm{P}(\mathbf{z})$
- If $\mathbf{x}_{\mathrm{new}} = \mathbf{x} \rightarrow \mathbf{x}_{\mathrm{new}} \sim \mathrm{U}(\mathrm{R}^*(\mathbf{x}))$

Optional Hybridization

- Way how to improve heuristics
- Evaluation of $\mathrm{f}(\mathbf{x})$ in better solution whenever possible
- $\mathrm{H} : \mathcal{D} \to \mathcal{D}$
- Overwrites solution $\mathbf{x}$ by

$$\mathbf{x}_{\mathrm{hyb}} = \mathrm{H}(\mathbf{x}) \tag{18}$$

  where $\mathrm{f}(\mathbf{x}_{\mathrm{hyb}}) \leq \mathrm{f}(\mathbf{x})$
- Examples
  - cluster rearrangement via k-means strategy in the cluster analysis
  - local search with neighborhood radius $r = 1$ in any optimization task

Competitive Approach

- Mutation operators are randomly selected in every step according to their efficiency
- When $\mathrm{f}(\mathbf{x}_{\mathrm{new}}) < \mathrm{f}(\mathbf{x})$ then the mutation score is incremented by one
- Probability of mutation selection is proportional to the score
- Portfolio of T_{mut}:
  - $T_{\mathrm{mut}} \to 0+$ (Random Descent)
  - $T_{\mathrm{mut}} = 0.01, \ldots, 10$
  - $T_{\mathrm{mut}} \to +\infty$ (Random Shooting)

Algorithm of RMFIF

1: Set counters $\mathbf{n} = n_0 \mathbf{1} \in \mathbb{N}^H$, and mutation portfolio
2: Initialize population $\mathbf{P}$ of size N by uniform sampling from $\mathcal{D}$
3: **while** $f_{\mathrm{best}} > f^*$ and $neval < N_{\mathrm{max}}$ **do**
    % Single step...
4: **end while**

Single step of RMFIF algorithm

Using systematic selection strategy find $\mathbf{x}_k$
Generate random index j according to mutation probabilities $p_j = n_j / \ ||\mathbf{n}||_1$
Perform mutation $\mathbf{x}_{\mathrm{new}} = \mathrm{M}_j(\mathbf{x}_k)$
Evaluate $f_{\mathrm{new}} = \mathrm{f}(\mathbf{x}_{\mathrm{new}})$
if $f_{\mathrm{new}} < f_k$ **then**
    Update $n_j = n_j + 1$, $\mathbf{x}_k = \mathbf{x}_{\mathrm{new}}$, $f_k = f_{\mathrm{new}}$
    if $\min_{i=1,\dots,H} p_i = \frac{\min n_i}{||\mathbf{n}||_1} < \frac{\delta}{H}$ **then**
        Reset counters as $\mathbf{n} = n_0\mathbf{1}$
    end if
end if

Part III: The Best Submodel Selection

- Linear combination of explanatory variables $\mathbf{q}$
- Single response variable $\mathbf{y}^*$
- Logit, Probit, Poisson regression, linear regression, etc.
- Submodel selection via binary vector $\mathbf{b} \in \{0, 1\}^n$
- Binary optimization task which finds the best submodel of size $p = \sum_{j=1}^{n} b_j$
- Details in:
  - M. Mojzeš, J. Kukal and H. Lauschmann. A Comparison of Bayesian and Another Approaches in Estimation of Fatigue Crack Growth Rate From 2D Textural Features. *Journal of Theoretical and Applied Mechanics*, Accepted for publication
  - E. Bolcekova, J. Kukal, S. Ostry, M. Mojzes, V. Q. Tran, P. Kulistak and R. Rusina. Cognitive impairment in isolated cerebellar lesions: A logit model based on neuropsychological testing. *Cerebellum & Ataxias*, Accepted for publication

Generalized Linear Model (GLM)

- Probability density function (PDF) of a submodel:
  $f = \mathrm{f}\left(\sum_{j=0}^{p} a_j q_j, y^* \right)$
- Negative log likelihood is subject of inner continuous optimization:

$$\Phi(\mathbf{a}) = -\sum_{i=1}^{m} \ln \mathrm{f}\left(\sum_{j=0}^{p} a_j q_{i,j}, y_i^* \right) \qquad (19)$$

- Objective function for outer binary optimization:

$$\psi(\mathbf{b}) = \log_{10} p_{\mathrm{value}} + \lambda_1 \cdot \sum_{k=1}^{p} \mathrm{I}(|z_k| \leq \theta) + \lambda_2 \cdot \mathrm{I}(p > p_{\mathrm{max}}) = \min \qquad (20)$$

where $\lambda_1, \lambda_2 > 0$ are penalization gains, $\mathbf{b} \in \{0,1\}^n$ is selection vector, $p = \sum_{j=1}^{n} b_j$ is parameter number, $z_k = a_k^{\mathrm{opt}} / s_k$ is z-score of k-th parameter, θ is threshold ($\theta = 1.960$) and p_{max} is maximum parameter number

- Defined by

$$\text{prob}(y^* = 1 \mid \mathbf{q}) = \left(1 + \exp\left(-\sum_{j=0}^{p} a_j q_j\right)\right)^{-1} \tag{21}$$

where $y^* \in \{0, 1\}$, $t \in \mathbb{R}$

- PDF:

$$\mathrm{f}(t, y^*) = y^* \cdot (1 + \exp(-t))^{-1} + (1 - y^*) \cdot (1 - (1 + \exp(-t))^{-1}) \tag{22}$$

- Negative log likelihood:

$$-\ln \mathrm{f} = y^* \cdot \ln(1 + \exp(-t)) + (1 - y^*) \cdot (\ln(1 + \exp(-t)) + t) \tag{23}$$

- Defined by

$$y^* \sim \mathrm{Poiss}(\sum_{j=0}^{p} a_j q_j) = \mathrm{Poiss}(t) \tag{24}$$

where $y^* \in \mathbb{N}_0,\ t > 0$

- PDF:

$$\mathrm{f}(t, y^*) = \frac{t^{y^*}}{(y^*)!}\mathrm{e}^{-t} \tag{25}$$

- Negative log likelihood:

$$-\ln \mathrm{f} = t - y^* \ln t + \ln(y^*!) \tag{26}$$

Part IV:

- Experimental Results

Time Complexity Measure Comparison

Table 1: Optimal FSA time complexity measure values for different T_0 on Integer Interference Problem

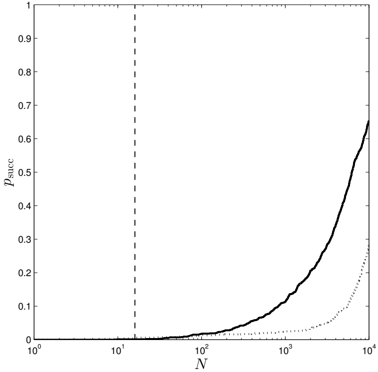
T_0	$n^*_{0.25}$	$n^*_{0.50}$	$n^*_{0.75}$	$n^*_{0.90}$	$n^*_{0.95}$	$n^*_{0.99}$	$\mathrm{E}\, n^*$	$\mathrm{E}\, n^*(0.1)$	$\mathrm{E} \ln n^*$	$\mathrm{E}(n^*)^{-1} \cdot 10^6$	*FEO*
0.001	7 596	18 734	37 506	61 668	80 064	122 423	26 837	21 136	9.60	**3 397**	**1 500**
0.01	9 647	23 325	46 650	78 177	101 401	154 262	34 428	28 913	9.95	218	22 893
0.1	11 315	26 067	49 406	79 009	107 289	158 580	36 616	29 241	10.00	236	23 606
1	6 521	15 572	30 741	50 556	65 520	100 820	22 142	19 389	9.48	283	15 148
10	9 099	21 761	39 416	64 696	80 203	121 128	29 410	25 601	9.83	325	20 375
100	**3 886**	**9 246**	**18 492**	**30 820**	**40 066**	**61 506**	**13 400**	**12 903**	**8.95**	503	10 225
1000	5 161	12 053	24 106	40 167	52 220	80 117	17 522	16 021	9.23	369	11 358

Time Complexity Measure Comparison

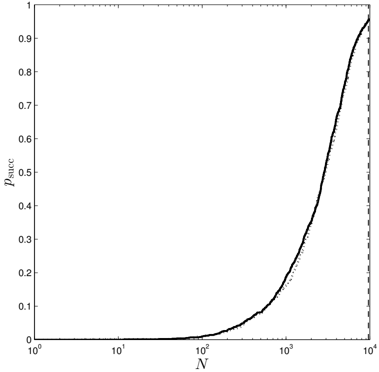
Table 2: Optimal FSA cut-off time for different T_0 on Integer Interference Problem

T_0	$n^*_{0.25}$	$n^*_{0.50}$	$n^*_{0.75}$	$n^*_{0.90}$	$n^*_{0.95}$	$n^*_{0.99}$	$\mathrm{E}\, n^*$	$\mathrm{E}\, n^*(0.1)$	$\mathrm{E} \ln n^*$	$\mathrm{E}(n^*)^{-1}$	*FEO*
0.001	2 527	2 224	2 527	2 224	2 224	2 224	2 224	2 224	2 224	**15**	**15**
0.01	4 665	4 665	4 665	4 665	4 665	4 665	4 665	4 665	4 665	5 860	7 274
0.1	11 315	13 639	38 988	38 988	38 988	38 988	38 988	38 988	13 639	1 416	1 416
1	1 820	1 820	1 820	1 820	1 820	1 820	1 820	1 820	1 820	1 621	1 820
10	9 099	13 757	39 416	40 503	40 503	40 503	40 503	40 503	40 503	428	428
100	**134**	**134**	**134**	**134**	**134**	**134**	**134**	**134**	**134**	137	275
1000	1 113	709	709	709	709	709	709	709	709	550	1 113

Optimal Restarting Period Using $\mathrm{E}\, n^*$



(a) $T = 0.283$
$\mathrm{E}\, n^*_{10{,}000} = 6,091$
$\mathrm{E}\, n^*_{16} = 4,039$



(b) $T = 0.400$
$\mathrm{E}\, n^*_{10{,}000} = 3,259$
$\mathrm{E}\, n^*_{9{,}625} = 3,130$

Figure 1: Reliability of two FSA instances on Clerc's Zebra3 for $d = 30$

Tuning of RMFIF via T_{MFT} and N

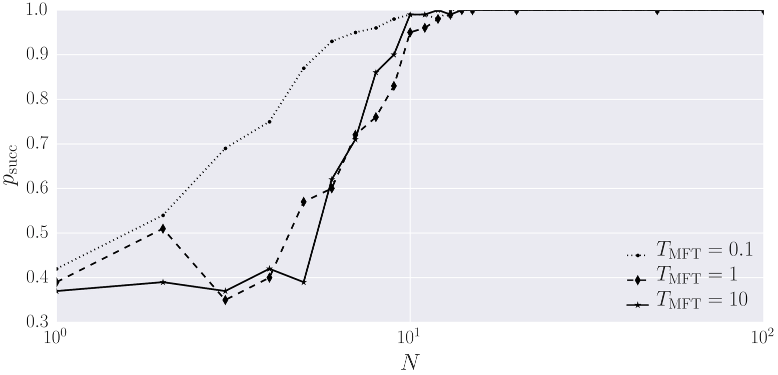


Figure 2: Reliability on the coloring problem as function of T_{MFT} and N (population size)

Tuning of RMFIF via T_{MFT} and N

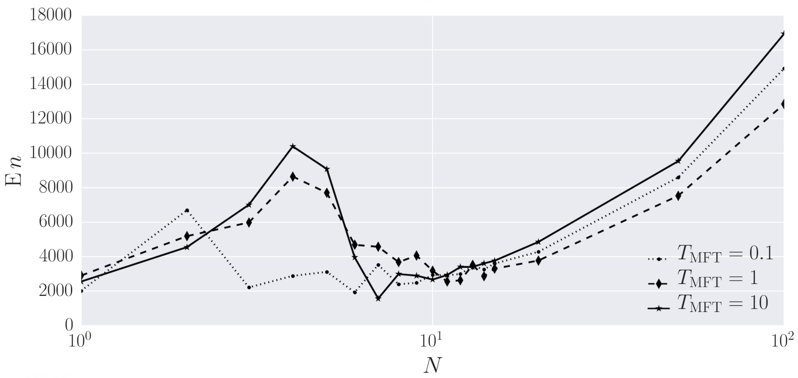


Figure 3: Mean number of evaluations on the coloring problem as function of T_{MFT} and N (population size)

Tuning of RMFIF via T_{MFT} and N

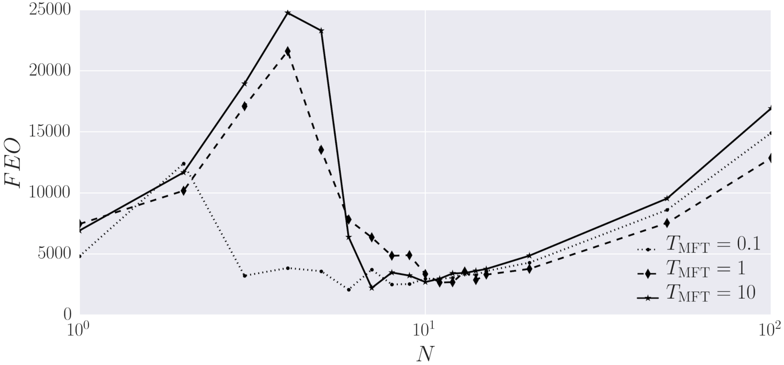


Figure 4: Feoktistov criterion on the coloring problem as function of T_{MFT} and N (population size)

RMFIF Efficiency via Traditional Measures

Table 3: Performance measures and parameter setting of the best performing RMFIF and referential heuristics

Task	Heuristic	Parameters	$\mathrm{E}\, n$	$\sqrt{\mathrm{D}\, n}$	p_{succ} [%]	*FEO*
Coloring $d = 49$	FSA	$\sigma = 0.25$	3,245	2,584	100	3,245
	GA	$\sigma = 3,\ N = 3$	12,166	13,277	100	12,166
	ILS	–	2,532	1,868	100	2,532
	RMFIF	$T_{\mathrm{MFT}} = 0.1,\ N = 6$	1,933	881	93	2,079
Hilbert $d = 16$	FSA	$\sigma = 1$	4,540	2,892	39	11,641
	GA	$\sigma = 10,\ N = 10$	4,099	2,760	32	12,810
	ILS	–	5,247	3,297	5	104,944
	RMFIF	$T_{\mathrm{MFT}} = 1,\ N = 15$	3,791	2,833	85	4,460
Zebra-3 $d = 15$	FSA	$\sigma = 0.1$	4,240	2,981	64	6,624
	GA	$\sigma = 5,\ N = 5$	4,262	2,732	71	6,003
	ILS	–	5,257	2,512	29	18,128
	RMFIF	$T_{\mathrm{MFT}} = 10,\ N = 15$	728	230	96	758

Table 4: Optimal submodel quality and features using hybrid heuristic

				Term								
Method	*CRIT*	R	k_{opt}	f_u	$f_u^{1/2}$	f_u^2	f_u^{-1}	$\log_{10} f_u$	F_{uv}	$F_{uv}^{1/2}$	F_{uv}^2	F_{uv}^{-1}
R^2 test	-117.92	0.9909	27	0	0	0	0	0	7	4	11	5
Wald test	-100.29	0.9839	15	0	0	0	1	0	1	3	7	3
WIC	-1,099.00	0.9993	88	1	1	4	1	1	14	19	22	25
AIC	-927.03	0.9992	82	0	1	2	1	0	20	19	18	21
BIC	-626.86	0.9871	20	0	0	1	0	1	1	2	13	2

Poisson Regression Submodel

Table 5: Optimal submodel of Poisson regression, $d = 14$

Heuristic	f_{best}	p_{best}	$\mathrm{E}\, n$	$\sqrt{\mathrm{D}\, n}$	$p_{\mathrm{succ}}[\%]$	*FEO*
Bottom-up	-1.9636	1	28	0	0	
Top-down	-1.5371	2	17	0	0	
$\mathcal{R}^*$ steepest descent	-1.9636	1	29	0	0	
RS	-2.2054	3	36	7	33	**109**
RMFIF	-2.2054	3	176	137	**98**	180

Logistic Regression Submodel in Biomedicine

- Explanatory variables: 30
- Observations: 50 (25 lesions and 25 controls)
- RMFIF parameters: $T_{\mathrm{MFT}} = 1$, $N = 100$, $\alpha = 1$, $\delta = 0.2$, $\sigma = 10^{-3}$, and mutation family based on $T_{\mathrm{mut}} \in \{0.001, 0.1, 0.2, 0.5, 1, 2, 5, 10, 1000\}$

Table 6: The best submodel of the test battery

Explanatory variable	Coefficient	Standard error	p_{value}
const	-14.4480	5.2470	0.0059
Education level	0.3970	0.2020	0.0493
CVLT-9 1-5 errors	5.8210	2.0220	0.0040
PST D time	0.4520	0.1550	0.0035

- Thank you for your attention