Case Study: Constraint Handling in Evolutionary Optimization of Catalytic Materials

Martin Holeňa*
Institute of Computer Science
Pod vodárenskou věží 2
CZ-18200 Prague
Czech Republic
http://www.cs.cas.cz/∼martin
martin@cs.cas.cz

David Linke
Leibniz Institute for Catalysis
Albert-Einstein Str. 29
D-18059 Rostock
Germany
david.linke@catalysis.de

Lukáš Bajer
Institute of Computer Science
Pod vodárenskou věží 2
CZ-18200 Prague
Czech Republic
bajer@cs.cas.cz

ABSTRACT

The paper presents a case study in an industrially important application domain – the optimization of catalytic materials. Though evolutionary algorithms are the by far most frequent approach to optimization tasks in that domain, they are challenged by mixing continuous and discrete variables, and especially by a large number of constraints. The paper describes the various kinds of encountered constraints, and explains constraint handling in GENACAT, one of evolutionary optimization systems developed specifically for catalyst optimization. In particular, it is shown that the interplay between cardinality constraints and linear equality and inequality constraints allows GENACAT to efficienlty determine the set of feasible solutions, and to split the original optimization task into a sequence of discrete and continuous optimization. Finally, the genetic operations employed in the discrete optimization are sketched, among which crossover is based on an assumption about the importance of the choice of sets of continuous variables in the cardinality constraints.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*Constrained optimization*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Algorithms

*corresponding author

GECCO'11, July 12–16, 2011, Dublin, Ireland

Keywords

evolutionary optimization, mixed optimization, equality constraints, inequality constraints, cardinality constraints

1. INTRODUCTION

In chemical engineering, much effort is devoted to increasing the performance of industrially important reactions, i.e., to achieving a higher yield of the desired reaction products without higher material or energy costs. Over 90% of chemical processes use a catalyst to this end. Catalysts are materials that decrease the energy needed to activate a chemical reaction without being themselves consumed in it. They typically consist of several components with different purpose, which can be selected from among many substances. Chemical properties of those substances constrain the possible ratios of their proportions, but they still allow for an infinite number of catalyst compositions. Moreover, the catalyst can usually be prepared from the individual components in a number of ways, and the preparation method also influences its performance in the chemical process. Consequently, the search for new catalytic materials leading to optimal performance of a chemical reaction entails high-dimensional constrained optimization tasks. Their *objective functions* are black-box functions, with values *obtained empirically*. Commonly used smooth optimization methods are not convenient to this end. Indeed, the discernibillity of the obtained measurements is too low to allow obtaining sufficiently precise numerical estimates of gradients or second order derivatives of the empirical objective function. Therefore, *methods not requiring derivatives* have been employed to solve those optimization tasks - both deterministic ones, in particular the *simplex method*, and stochastic ones, such as *simulated annealing*, or *evolutionary algorithms* (EA) [1]. Evolutionary, especially *genetic algorithms* (GA) are encountered most frequently, but their application to this area is far from any standard methodology. Main difficulties on a way to such a methodology are *mixed optimization* with respect to continuous and discrete variables, and *constraints*.

This paper deals with the latter difficulty. It describes how the constraints faced by the optimization of catalytic materials are tackled in the evolutionary optimization system GENACAT, developed in recent years at the Leibniz Institute for Catalysis in Rostock, in collaboration with the Institute for Computer Science in Prague. The overall func-

tionality of the system has been outlined, from the point of view of the application domain, in the Journal of Chemical Information and Modeling [12]. The present paper, on the other hand, deals solely with the method of constraint handling adopted in GENACAT, and with its incorporation into evolutionary optimization.

Due to the specific meaning conveyed by coordinates of points in the input space of the objective function, it is quite difficult to use general EA software, such as the Global Optimization Toolbox of Matlab [17]. Indeed, such general software only optimizes functions with input spaces of low-level data types, such as vectors of real numbers and bit-strings. And encoding the qualitative and quantitative compositions of catalytic materials, and their preparation and reaction conditions with low-level data types is a tedious and error-prone activity. Furthermore, it requires a great deal of mathematical erudition. Therefore, it is not surprising that general EA software has been used rarely in catalysis: at the very beginning [18, 19], in a recent application of genetic programming [2], and in most cases when multiobjective genetic optimization was performed [4, 9], especially if the popular non-dominated sorting genetic algorithm was used [5, 15]. Apart from those exceptions, the application of EA in this area took the route of developing specific algorithms for the optimization of catalytic materials. Although such algorithms have been developed in various institutions throughout the world [16, 20, 25], main contributions to their development came from the Institute for Catalytic and Environmental Research in France [3, 6, 21], the Institute of Chemical Technology in Spain [7, 8, 23, 24], and the Leibniz Institute for Catalysis (LIKAT) in Germany [10, 12, 22, 26]. The most recent contribution of the last mentioned institution is the above mentioned system GENACAT.

In the next section, the various kinds of constraints encountered in the optimization of catalytic materials are surveyed. The approach to constraint handling in GENACAT is presented in Section 3. In particular, 3.1 explains how all feasible solutions are determined, and 3.2 how sets of feasible solutions are evolved.

2. CONSTRAINTS FACED IN CATALYST OPTIMIZATION

In the optimization of catalytic materials, the objective function (in evolutionary optimization called fitness) is some empirical performance measure of the catalytic material, most frequently yield of (some of) the reaction product(s). The individual coordinates of points in the input space of the objective function typically convey some of the following meanings [1, 11]:

- **(i)** *Qualitative composition* of the catalytic material, i.e., of which components it consists, and what is its support.
- **(ii)** *Quantitative composition* of the catalytic material, i.e., the fractions of the various components mentioned in (i).
- **(iii)** *Preparation* of the catalytic material, its individual steps and their quantitative characterizations, such as temperatures or durations.
- **(iv)** *Reaction conditions* of the catalyzed reaction.

The random variables corresponding to (i) are always discrete and finitely-valued, the random variables corresponding to (ii) are always continuous, whereas the random variables corresponding to (iii)–(iv) can be of both kinds. Denoting d the vector of input coordinates corresponding to discrete random variables and x the vector of input coordinates corresponding to continuous random variables, the considered optimization task can be formulated as:

$$\text{maximize } f(d,x) \text{ subject to explicit constraints } c_1,\ldots,c_{n_c}, \tag{1}$$

An example of a real optimization task of that kind is given in Figure 1.

The term *explicit constraints* employed for the constraints $c_1,\ldots,c_{n_c}$ in (1) refers to the fact that the optimization is constrained also implicitely, through interval constraints on the value sets of the involved continuous random variables are. Taking into account different kinds of encountered explicit constraints, we differentiate altogether 5 kinds of constraints in the optimization of catalytic materials:

1. Bounded-interval constraints on the value sets of the continuous random variables. Most frequently, the lower bound of the interval is 0, thus $\mathrm{Val}(X_i) = [0,c]$ with $c > 0$ (in the example in Figure 1, this is the case for the variables $X_1 - X_{41}$, i.e., for 41 among the 42 involved continuous variables). In the sequel, the set of involved continuous random variables with this property will be denoted $\mathcal{X}_0$.

2. Equality or inequality constraints interconnecting the components of the vector d, which are realizations of discrete random variables $D_1,\ldots,D_{n_d}$ (in Figure 1, there is one such constraint: c_{97}). They delimit, in the cartesian product $\mathrm{Val}(D_1)\times\cdots\times\mathrm{Val}(D_{n_d})$ of the finite value sets of $D_1,\ldots,D_{n_d}$, a subset $\mathcal{D}$ of admissible values of d.

3. For each fixed $d \in \mathcal{D}$, systems of linear equations or inequalities for the vector x,

$$A_{=}^{(d)}x = b_{=}^{(d)}, \tag{2}$$

$$A_{\leq}^{(d)}x \leq b_{\leq}^{(d)}, \tag{3}$$

   where $A_{=}^{(d)}$, $A_{\leq}^{(d)}$ are matrices with as many columns as is the dimension of x (in Figure 1, (2) corresponds to the constraints $c_{92} - c_{95}$, whereas (3) corresponds to the constraint c_{96}). Observe that for a fixed d, the components of d occurring in (2) and (3) are constants (cf. c_{92} in Figure 1).

   We would like to emphasize that the linear equations and inequalities in (2)–(3) actually allow to describe constraints

$$g_{=}^{(d)}(x) = 0 \text{ or } g_{\leq}^{(d)}(x) \leq 0, \tag{4}$$

   where $g_{=}^{(d)}$ and $g_{\leq}^{(d)}$ are real functions that are either continuous or have a finite number of discontinuities of the first kind, i.e., they can be made left- and right-continuous in the discontinuities (this is always true for nonlinear constraints encountered in catalysis). Indeed, such functions can be aproximated with piecewise-linear functions, and the linear piece containing x can be indicated with the value of an additional discrete random variable. This leads to increasing the dimension of d and to a different set $\mathcal{D}$, whereas the

Fitness: Y (product yield)
Continuous variables:
X_i: proportion of the i-th component from the components pool available for the catalytic material, $i = 1, \ldots, 37$, $\mathrm{Val}(X_i) = [0, 0.1]$ for $i = 1, \ldots, 22$, $\mathrm{Val}(X_i) = [0, 0.003]$ for $i = 23, \ldots, 37$;
X_{38}: overall proportion of components belonging to precious metals, $\mathrm{Val}(X_{38}) = [0, 0.003]$;
X_{39}: overall proportion of components belonging to alkaline earth metals or lanthanoids, $\mathrm{Val}(X_{39}) = [0, 0.05]$;
X_{40}: proportion of the lower valence element in a fixed pair of alkaline earth metals or lanthanoids, $\mathrm{Val}(X_{40}) = [0, 0.01]$;
X_{41}: proportion of the higher valence element in a fixed pair of alkaline earth metals or lanthanoids, $\mathrm{Val}(X_{41}) = [0, 0.05]$;
X_{42}: overall proportion of components not belonging to precious or to alkaline earth metals or lanthanoids, $\mathrm{Val}(X_{42}) = = [0.003, 0.05]$;
Discrete variables:
D_1: choice of a material serving as support of the catalyst, $\mathrm{Val}(D_1) =$ {material1, material2};
D_2: proportion of support; $\mathrm{Val}(D_2) =$ {0.95,0.99};
D_3: choice of a fixed pair of alkaline earth metals or lanthanoids, $\mathrm{Val}(D_3) =$ {(i-th component, i'-thcomponent): $(29 \leq i \leq 31$ & $32 \leq i' \leq 33) \vee (32 \leq i \leq 33$ & $34 \leq i' \leq 37)$};
D_4: number of included components belonging to precious metals, $\mathrm{Val}(D_4) = \{0, 1\}$;
D_5: number of included components belonging to alkaline earth metals or lanthanoids, $\mathrm{Val}(D_5) = \{0, 1, 2\}$;
D_6: number of included fixed pairs of alkaline earth metals or lanthanoids, $\mathrm{Val}(D_6) = \{0, 1\}$;
D_7: number of included components belonging neither to precious metals, nor to alkaline earth metals or lanthanoids, $\mathrm{Val}(D_7) = \{1, 2, 3, 4\}$;
D_8: overall number of all included components, $\mathrm{Val}(D_8) = \{2, 3, 4\}$;
Constraints:
c_i: probability distribution of X_i on $[0.003, 0.1]$ is uniform, $i = 1, \ldots, 22$;
c_i: probability distribution of X_i on $(0, 0.003]$ is uniform, $i = 23, \ldots, 37$;
c_{38}: joint probability distribution of (X_{40}, X_{41}) on $\{(x, x') : 0 \leq x \leq 0.01$ & $0 \leq x' \leq 0.05$ & $20x \leq x' \leq 50x\}$ is uniform;
c_i: $P(0 < X_{i-38} < 0.003) = 0, i = 39, \ldots, 60$;
c_i:$P(X_1 > 0.03) = 3 * P(X_{i-59} > 0.03)$, $i = 61, \ldots, 81$;
c_j: probability distribution of D_{j-60} on $\mathrm{Val}(D_{j-60})$ is uniform, $j = 82, \ldots, 84$;
c_{85}: probability distribution of D_4 on $\{0, 1\}$ is $(0.8, 0.2)$;
c_{86}: joint probability distribution of (D_5, D_6) on $\mathrm{Val}(D_5) \times \mathrm{Val}(D_6) = \{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1)\}$ is $(\frac{1}{3}, \frac{1}{3}, \frac{2}{9}, \frac{1}{9}, 0)$;
c_{87}: probability distribution of D_8 on $\{2, 3, 4\}$ is $(0.45, 0.45, 0.1)$;
c_{88}: $|\{i : 23 \leq i \leq 28$ & $x_i > 0\}| = d_4$;
c_{89}: $|\{i : 29 \leq i \leq 37$ & $x_i > 0\}| = d_5$;
c_{90}: $|\{i : i = 40$ & $x_i > 0\}| = d_6$;
c_{91}: $|\{i : 1 \leq i \leq 22$ & $x_i > 0\}| = d_7$;
c_{92}: $d_2 + x_{38} + x_{39} + x_{42} = 1$;
c_{93}: $x_1 + x_2 + \cdots + x_{22} = x_{42}$;
c_{94}: $x_{23} + x_{24} + \cdots + x_{28} = x_{38}$;
c_{95}: $x_{29} + x_{30} + \cdots + x_{37} + x_{40} + x_{41} = x_{39}$;
c_{96}: $20x_{40} \leq x_{41} \leq 50x_{40}$;
c_{97}: $d_4 + d_5 + d_6 + d_7 = d_8$.

Figure 1: Example of a real task in the optimization of catalytic materials

equations and inequalities (4) are replaced with additional rows in (2)–(3). Due to the comparatively low discernibility of catalytic measurements, the piecewise-linear approximation of the functions $g_{=}^{(d)}$, resp. $g_{\leq}^{(d)}$, will even for a small number of linear pieces typically lead to an approximation error that is lower than the measurement error.

4. For some mutually exclusive sets of variables $S \subset \mathcal{X}_0$, constraints on the cardinality $|S_{>0}|$ of the set

$$S_{>0} = \{X \in S : X > 0\}, \tag{5}$$

which take the form of an equality

$$(\forall d \in \mathcal{D})|S_{>0}| = s(d), \tag{6}$$

with an S-specific fuction $s : \mathcal{D} \to \mathcal{N}_0$ (in Figure 1: the constraints $c_{88} - c_{91}$). Consequently, for each fixed $d \in \mathcal{D}$, the cardinality of S_0 is constant. Observe that (6) substantially simplifies solving the system (2)–(3) of equations and inequalities: It allows to solve $\binom{|S|}{s(d)}$ lower-dimensional systems instead, corresponding to the $\binom{|S|}{s(d)}$ possible choices of S_0. In each such system the $|S| - s(d)$ columns corresponding to the variables from $S \setminus S_0$ are left out from the matrix $A_{=}^{(d)}$ in (2) and the matrix $A_{\leq}^{(d)}$ in (3). To illustrate how large simplification this may entail, consider the constraint c_{91} in Figure 1. Alone due to this single constraint, 18 to 21 columns are left out from among the 42 columns of the matrix $A_{=}^{(d)}$, corresponding to the constraints $c_{92} - c_{95}$, and the matrix $A_{\leq}^{(d)}$, corresponding to the constraint c_{96} (18 columns if $d_7 = 4$,...,21 columns if $d_7 = 1$).

5. Required distributions of the involved discrete and continuous random variables (in Figure 1: the constraints $c_1 - c_{87}$). By default, the distribution of each random variable, no matter whether discrete or continuous, is required to be uniform. However, this can be explictely changed. The distributions influence how the 1st generation of the evolutionary algorithm looks out, and what is the result of mutation. However, it is important to realize that they are constraints on the underlying random variables, not on their realizations, even not on populations of their realizations, like those produced in evolutionary algorithms. Hence, a vector (x, d) can never be unfeasible with respect to required distributions, nor can be a population of such vectors, irrespectively of how the empirical distributions of their coordinates differ from the required distributions of the underlying random variables.

3. CONSTRAINT HANDLING IN THE SYSTEM "GENACAT"

In GENACAT, a specific GA precisely *tailored to the optimization task being solved* is generated by a program generator, based on a user specification of the task in a *catalyst description language* (CDL). The implementation of the program generator, the CDL-language, as well as the creating

and processing of CDL-descriptions have been presented in [12]. Here on the other hand, we explain the method used in GENACAT to solve the mixed constrained optimization task (1). It is based on the fact that for each fixed $d \in \mathcal{D}$ and each choice of S_0 in any cardinality constraint (6), the set of feasible solutions to (2)–(3) is a polyhedron, determined by some matrix A_P and vector b_P

$$P = \{\xi : A_P\xi \leq b_P\}. \tag{7}$$

Each such polyhedron can be empty, and each has its specific dimension, ranging between 1 (closed interval) and the number of input coordinates corresponding to continuous random variables.

3.1 Finding Feasible Solution Polyhedra

If a solution polyhedron is described with (7), then its feasibility (i.e., non-emptiness) is invariant with respect to any permutation of columns of A_P, and to any permutation of rows of $(A_P b_P)$. Moreover, the relation $\approx$ defined

$$\begin{aligned} P \approx Q \text{ iff } (A_Q b_Q) \text{ can be obtained from } (A_P b_P) \text{ by} \\ \text{a permutation of columns of } A_P, \text{ followed by} \\ \text{a permutation of rows of the result and of } b_P \end{aligned} \tag{8}$$

is an equivalence, partitioning the set of polyhedra into disjoint classes. This property plays a key role in GENACAT because only one representative from each class needs to be checked for non-emptiness.

In Figure 2, the difference between the overall number of polyhedra and the number of their equivalence classes, as well as between the number of feasible polyhedra and the number of equivalence classes with feasible polyhedra is illustrated on 5 example catalyst optimization tasks that have been solved by the system GENACAT. Among them, the task nr. 5 is the one presented above in Figure 1.

3.2 Evolution of Feasible Polyhedra

On the set of feasible polyhedra, discrete genetic optimization is performed, using operations selection, mutation and crossover developed specifically to this end. Each of the polyhedra forming the population obtained in this way contains a subpopulation of combinations of values of continuous variables found by linearly constrained continuous genetic optimization (we used its implementation in the Global Optimization Toolbox of Matlab [17] to this end). The union of all such subpopulations combined with the combinations of values of discrete variables corresponding to the respective polyhedra, form together the final population of solutions to the optimization task. The specific genetic operations employed in the discrete optimization are defined as follows:

Selection is in the first generation uniform, in subsequent generations proportional to the importance of the polyhedron due to points from earlier generations that it contains. As a measure of that importance, the difference between the fitness of a point and the minimal fitness encountered in previous generations is taken, summed over points with the combination of values of the discrete variables corresponding to the polyhedron.

Mutation consists in replacing an existing polyhedron with a uniformly selected nonempty one. The values of continuous variables forming a point in that polyhedron are again obtained by linearly constrained continuous genetic optimization. If the mutation rate is μ, then a proportion μ of the population is selected in this way, and the proportion $1 - \mu$ is selected using the above proportional selection.

	Example tasks				
	1	2	3	4	5
continuous variables	11	14	28	29	42
input space constraints	23	32	51	53	97
	Polyhedra				
all	22275	1512	$5.1 \cdot 10^6$	$6.1 \cdot 10^6$	$583 \cdot 10^6$
feasible	22110	1000	224620	171140	282810
	Classes of polyhedra				
all	6	36	1323	756	480
feasible	5	26	312	178	60

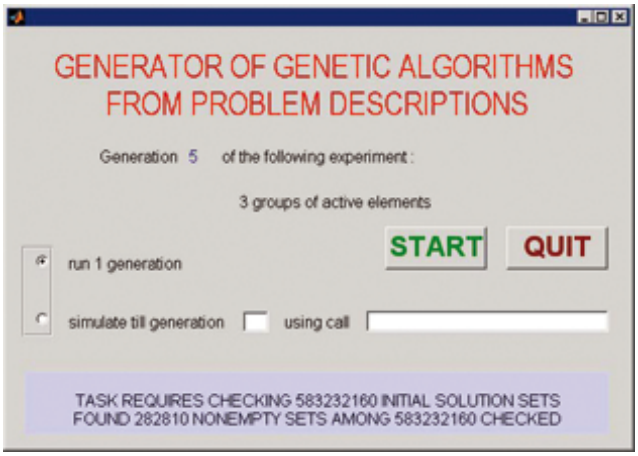

Figure 2: Top: comparison between the number of (all and feasible) polyhedra and their equivalence classes for 5 example catalyst optimization tasks solved by the system GENACAT, including the task presented in Figure 1 (task 5). Bottom: End-users are not confronted with the concept of equivalence classes, that is why the graphical interface controlling the evolutionary optimization by GENACAT reports only the numbers of solution polyhedra

Crossover relies on the fact that a solution polyhedron P is determined on the one hand by the choice of the vector $d = d^P$ of values of discrete variables, on the other hand by the choice of the set $S_0 = S_0^P$ of continuous variables in every cardinality constraint (6). We assume that for the optimization of catalytic materials, the choice of the sets of continuous variables is more important, and we suggest a crossover operation that always exchanges exactly one of the continuous variables corresponding to the parent polyhedra, and attempts to include as high agreement with their vectors of values of discrete variables as possible. Formally, denote m_p the number of columns of A_P, and let $x_{P,1}, \ldots, x_{P,m_P}$ be the continuous variables corresponding to those columns. If the crossover rate is λ, then for each pair P and P' of solution polyhedra selected using the proportional selection, a set of recombination offsprings is formed with probability λ, in the following way:

(i) The set of candidate offsprings of P and P' is defined

by

$$\begin{aligned}\mathcal{C}(P,P') = &\{Q - \text{ solution polyhedron } : Q \neq \emptyset \;\&\\ &\&\ \{x_{Q,1},\ldots,x_{Q,m_Q}\} \subset \{x_{P,1},\ldots,x_{P,m_P}\} \cup\\ &\cup\ \{x_{P',1},\ldots,x_{P',m_{P'}}\}\ \&\\ &\&\ [(m_Q = m_P\ \&\ |\{x_{Q,1},\ldots,x_{Q,m_Q}\} \cap\\ &\cap\ \{x_{P,1},\ldots,x_{P,m_P}\}| = m_Q - 1) \vee\\ &\vee\ (m_Q = m_{P'}\ \&\ |\{x_{Q,1},\ldots,x_{Q,m_Q}\} \cap\\ &\cap\ \{x_{P',1},\ldots,x_{P',m_{P'}}\}| = m_Q - 1)]\}. \end{aligned} \tag{9}$$

(ii) For each $Q \in \mathcal{C}(P,P')$, the uncertainty index of Q is computed as

$$u(Q) = |\{j : 1 \leq j \leq n_d\ \&\ d_j^Q \notin \{d_j^P, d_j^{P'}\}\}|. \tag{10}$$

(iii) The final set of offsprings of P and P' is defined by

$$\mathcal{O}(P,P') = \{Q \in \mathcal{C}(P,P') : u(Q) = \min_{Q' \in \mathcal{C}(P,P')} u(Q')\}. \tag{11}$$

Due to the selection proportional to importance, the subpopulations of combinations of continuous variables in polyhedra with high importance tend to increase, whereas the subpopulations in polyhedra with low importance tend to decrease or to disappear. This is illustrated in Figure 3 for the example task introduced in Figure 1. For that task, the genertated GA was run with a population size 96, given by the number of available channels in the reactor in which the catalysts were tested. The evolution was finished after 7 generations, when the number of found catalysts with sufficiently high fitness (yield) was already satisfactory for the chemists, in view of the cost of the evaluation of another generation. Figure 3 shows the development of the distribution of subpopulation sizes during the $2^{\text{nd}} - 7^{\text{th}}$ generation.

4. CONCLUSIONS

This paper presented a case study in an industrially important application domain – the optimization of catalytic materials. The optimization tasks occurring in that domain are by far most frequently dealt with evolutionary algorithms, and constraints pose one of key difficulties connected with such tasks. The paper described the various kinds of encountered constraints, and then explained constrain handling in GENACAT, one of evolutionary optimization systems developed specifically for that application domain.

Most importantly, it was shown that the interplay between cardinality constraints and linear equality and inequality constraints allows GENACAT to efficienlty determine the set of feasible solutions, and to split the original optimization task into a sequence of discrete and continuous optimization. Moreover, the definition of the crossover operation employed in the discrete optimization is based on an assumption about the importance of the choice of the sets of continuous variables in the cardinality constraints.

At the same time, the case study reflects the specificity of evolutionary optimization in catalysis, which was already partially mentioned in the introduction. In particular:

1. GENACAT heavily relies on the specificity of constraints encountered in catalyst optimization, as well as on the specific meaning of the coordinates of points in the input space of the objective function. Therefore, we were not interested in comparing the constraint handling approach implemented in GENACAT with approaches implemented in publicly available systems, which do not take into account those specific features, and employ low-level coding of points in the input space.

2. The evaluation of objective functions used in catalysis is costly and time-consuming. Typically, the evaluation of one generation of catlytic materials needs several days to several weeks, and costs several to many thousands of euros. That is the reason why evolutionary optimization usually runs only for 5-10 generations, till the costs of evaluating a next generation of materials are not worth the expected improvement (cf. Figure 3)

Finally, we would like to emphasize that another system was used at LIKAT before GENACAT, in which constraints were handled insufficently [22, 26]. And that dealing with the various kinds of constraints described in Section 2 was actually one of two main objectives of GENACAT development, the other being its integration with surrogate modeling [13, 14].

5. ACKNOWLEDGMENTS

The research reported in this paper was supported by the German Federal Ministry of Education and Research, as well as by the Czech Science Foundation grants 201/08/0802 and P202/11/1368.

6. REFERENCES

[1] M. Baerns and M. Holeňa. *Combinatorial Development of Solid Catalytic Materials. Design of High-Throughput Experiments, Data Analysis, Data Mining*. World Scientific, Singapore, 2009.

[2] L. Baumes, A. Blanché, P. Serna, A. Tchougang, N. Lachiche, P. Collet, and A. Corma. Using genetic programming for advanced performance assessment of industrially relevant heterogeneous catalysts. *Materials and Manufacturing Processes*, 24:282–292, 2009.

[3] L. Baumes, P. Jouve, D. Farrusseng, M. Lengliz, N. Nicoloyannis, and C. Mirodatos. Dynamic control of the browsing-exploitation ratio for iterative optimisations. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 265–270. Springer Verlag, Berlin, 2003.

[4] C. Breuer, M. Lucas, F. Schuetze, and P. Claus. Implementation of the multi-channel monolith reactor in an optimisation procedure for heterogeneous oxidation catalysts based on genetic algorithms. *Combinatorial Chemistry and High Throughput Screening*, 10:59–70, 2007.

[5] S. Cheng, H. Chen, H. Chang, C. Chang, and Y. Chen. Multi-objective optimization for two catalytic membrane reactors - methanol synthesis and hydrogen production. *Chemical Engineering Science*, 63:1428–1437, 2008.

[6] F. Clerc, M. Lengliz, D. Farrusseng, C. Mirodatos, S. Pereira, and R. Rakotomata. Library design using genetic algorithms for catalysts discovery and

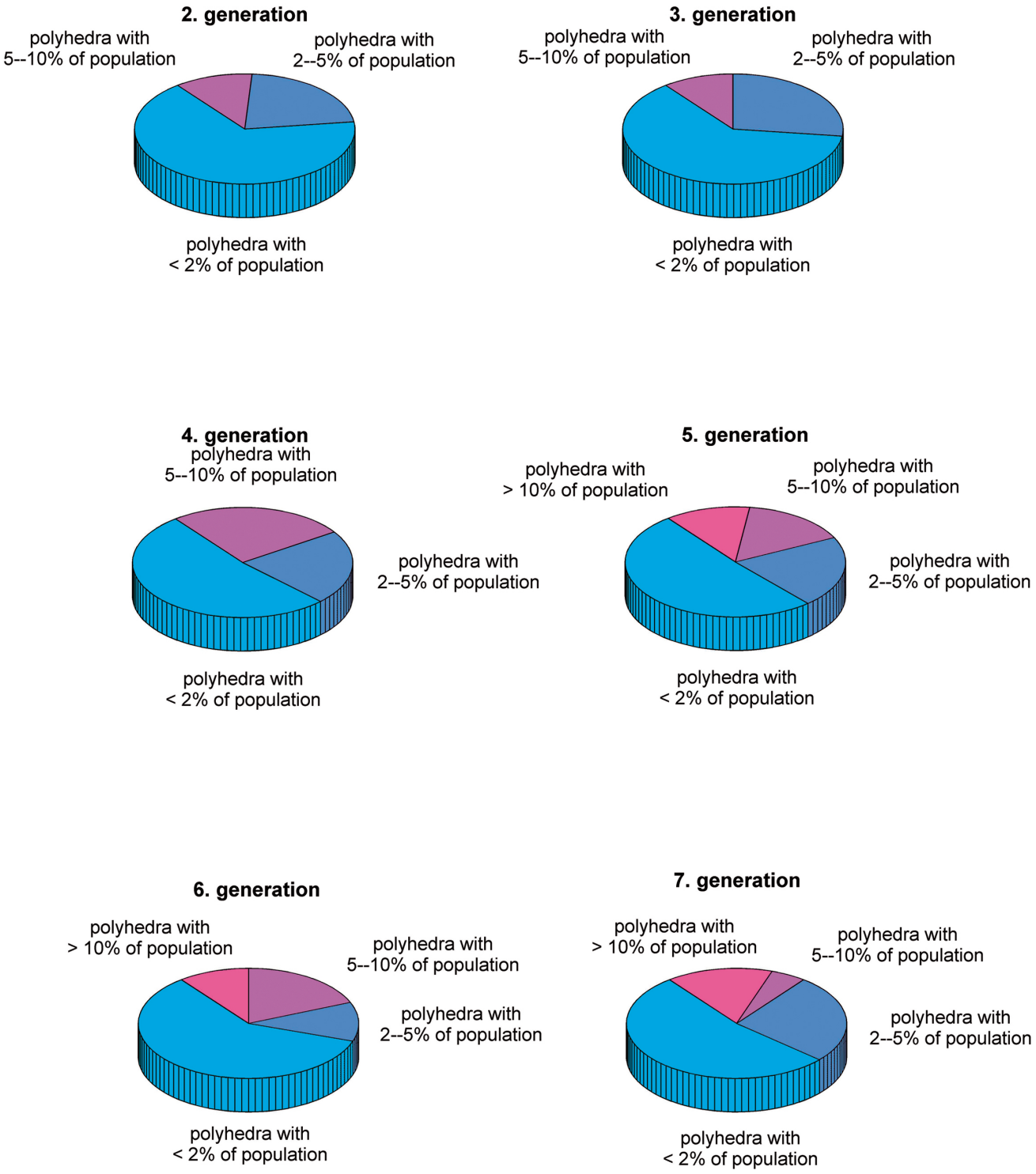

Figure 3: The distribution of sizes of the individual polyhedra during the $2^{\text{nd}} - 7^{\text{th}}$ generation of running the genertated GA for the optimization task in Figure 1

optimization. *Review of Scientific Instruments*, 76:062208, 2005.

[7] A. Corma, J. Serra, and A. Chica. Discovery of new paraffin isomerization catalysts based on SO_4^{2-}/ZrO_2 and WO_x/ZrO_2 applying combinatorial techniques. *Catalysis Today*, 81:495–506, 2003.

[8] A. Corma, J. Serra, P. Serna, S. Valero, E. Argente, and V. Botti. Optimisation of olefin epoxidation catalysts with the application of high-throughput and genetic algorithms assisted by artificial neural networks (softcomputing techniques). *Journal of Catalysis*, 229:513–524, 2005.

[9] O. Gobin and F. Schüth. On the suitability of different representations of solid catalysts for combinatoral library design by genetic algorithms. *Journal of Combinatorial Chemistry*, 10:835–846, 2008.

[10] G. Grubert, E. Kondratenko, S. Kolf, M. Baerns, P. van Geem, and R. Parton. Fundamental insights into the oxydative dehydrogenation of ethane to ethylene over catalytic materials discovered by an evolutionary approach. *Catalysis Today*, 81:337–345, 2003.

[11] M. Holeňa and M. Baerns. Computer-aided strategies for catalyst development. In G. Ertl, H. Knözinger, F. Schüth, and J. Eitkamp, editors, *Handbook of Heterogeneous Catalysis*, pages 66–81. Wiley-VCH, Weinheim, 2008.

[12] M. Holeňa, T. Cukic, U. Rodemerck, and D. Linke. Optimization of catalysts using specific, description based genetic algorithms. *Journal of Chemical Information and Modeling*, 48:274–282, 2008.

[13] M. Holeňa, D. Linke, and U. Rodemerck. Evolutionary optimization of catalysts assisted by neural-network learning. In *Simulated Evolution and Learning. Lecture Notes in Computer Science 6457*, pages 220–229. Springer Verlag, Berlin, 2010.

[14] M. Holeňa, D. Linke, and U. Rodemerck. Generator approach to evolutionary optimization of catalysts and its integration with surrogate modeling. *Catalysis Today*, 159:84–95, 2011.

[15] K. Keyvanloo and J. Towfighi. Comparing the catalytic performances of mixed molybdenum with cerium and lanthanide oxides supported on hzsm-5 by multiobjective optimization of catalyst compositions using nondominated sorting genetic algorithm. *Journal of Analytical and Applied Pyrolysis*, 88:140–148, 2010.

[16] G. Kirsten and W. Maier. Strategies for the discovery of new catalysts with combinatorial chemistry. *Applied Surface Science*, 223:87–101, 2004.

[17] The Mathworks, Inc. *Global Optimization Toolbox 3*, 2010.

[18] A. McLeod, M. Johnston, , and G. L.F. Development of a genetic algorithm for molecular scale catalyst design. *Journal of Catalysis*, 167:279–285, 1997.

[19] A. McLeod and G. L.F. Heterogeneous catalyst design using stochastic optimization algorithms. *Journal of Chemical Information and Computer Science*, 40:981–987, 2000.

[20] A. Ohrenberg, C. Törne, A. Schuppert, and B. Knab. Application of data mining and evolutionary optimization in catalyst discovery and high-throughput experimentation – techniques, strategies, and software. *QSAR and Combinatorial Science*, 24:29–37, 2005.

[21] R. Pereira, F. Clerc, D. Farrusseng, J. Waal, and T. Maschmeyer. Effect of genetic algorithm parameters on the optimization of heterogeneous catalysts. *QSAR and Combinatorial Science*, 24:45–57, 2005.

[22] U. Rodemerck, M. Baerns, and M. Holeňa. Application of a genetic algorithm and a neural network for the discovery and optimization of new solid catalytic materials. *Applied Surface Science*, 223:168–174, 2004.

[23] J. Serra, L. Baumes, M. Moliner, P. Serna, and A. Corma. Zeolite synthesis modelling with support vector machines: a combinatorial approach. *Combinatorial Chemistry and High Throughput Screening*, 10:13–24, 2007.

[24] J. Serra, A. Chica, and A. Corma. Development of a low temperature light paraffin isomerization catalysts with improved resistance to water and sulphur by combinatorial methods. *Applied Catalysis A: General*, 239:35–42, 2003.

[25] Y. Watanabe, Umegaki, T., Hashimoto, M., Omata, K., and M. Yamada. Optimization of Cu oxide catalysts for methanol synthesis by combinatorial tools using 96 well microplates, artificial neural network and genetic algorithm. *Catalysis Today*, 89:455–464, 2004.

[26] D. Wolf, O. Buyevskaya, and M. Baerns. An evolutionary approach in the combinatorial selection and optimization of catalytic materials. *Applied Catalyst A: General*, 200:63–77, 2000.