17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

An Algebraic Property Clustering Technique for Molecular Design

Fadwa T. Eljack, Charles C. Solvason, Mario R. Eden

Department of Chemical Engineering, 222 Ross Hall, Auburn University, Auburn, AL 36849, USA. Email: edenmar@auburn.edu

Abstract

The introduction of the property integration framework has enabled representation of processes and products from a properties perspective. The framework enables identification of the desired component properties by targeting optimum process performance without committing to any components during the solution step. To provide a unifying methodology for handling both process and molecular design problems, the property integration framework was extended to include Group Contribution Methods (GCM) to enable prediction of physical properties from structural information. In our previous work, the framework was limited to handling problems that could be adequately described using only three properties. In this contribution, an algebraic approach is presented that enables solution of problems requiring multiple properties, thus expanding the application range of the molecular property clustering technique.

Keywords: Property Integration, Molecular Design, Algebraic Approach

1. Molecular Property Clusters

The clustering technique utilizes property operators, which are functions of the original raw physical properties [1-4]. Although the operators themselves may be highly non-linear, they are tailored to possess linear mixing rules, e.g. density does not exhibit a linear mixing rule, however the reciprocal value of density follows a linear mixing rule [1-4]. Extending the original property

integration framework to include group contribution methods (GCM) for molecular design required the introduction of molecular property operators [4]. Fortunately, the equations employed in GCM are similar to the original property operator formulation, i.e. the properties are described by linear additive rules for the individual molecular fragments [5,6]. In Eq. (1), $\psi^M_j(P_j)$ is the molecular property operator of the *j*'th property. The RHS of the equation is always in the form of summation of the number of occurrences of each group (n_g) multiplied by the contribution to property *j* from group *g* (P_{jg}).

$$\psi_j^M(P_j) = \sum_{g=1}^{N_g} n_g \cdot P_{jg} \quad , \quad \Omega_j = \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \tag{1}$$

Next, the molecular property operators can be converted to clusters according to the procedures developed for the original property clusters [4]. The Augmented Property index AUP^M for each molecule is defined as the summation of all the *NP* dimensionless property operators (Ω_M), and finally the property cluster C_j is obtained as the ratio of Ω^M and AUP^M:

$$\Omega^M_{ji} = \frac{\psi^M_j(P_{ji})}{\psi^{ref}_j(P_{ji})} \quad , \quad AUP^M = \sum_{j=1}^{NP} \Omega^M_j \quad , \quad C_j = \frac{\Omega^M_j}{AUP^M} \tag{2}$$

2. Algebraic Property Clustering Approach

The clustering technique reduces the dimensionality of design problems, thus it is possible to visually identify the solutions, which is a significant advantage of this approach. The ability to synthesize molecules within the clustering domain is key to bridging the gap between process and molecular design, however until now it has been limited to problems using only three properties [4]. Here we will further exploit the advantages of the linear additive rules of the molecular operators to setup the design problem as a set of linear algebraic equations to synthesize molecular formulations, given a set of molecular building blocks (first order groups from GCM) represented by n_g and a set of property performance requirements/constraints that is described by:

$$P_{ij}^{lower} \le P_{ij} \le P_{ij}^{upper} \quad , \quad \Omega_j^{\min} \le \Omega_{ij} \le \Omega_j^{\max} \quad , \quad \Omega_j^{\min} \le \sum_{g=1}^{N_g} n_g \cdot \Omega_{jg} \le \Omega_j^{\max} \tag{3}$$

Because each property can be expressed in terms of two inequalities, each property can be combined with another property in two ways. In the original

visualization approach for the molecular design framework, the bounds on three properties can be represented by a set of six points [1-4]. Similarly, for systems made up of four properties, Ω_1-Ω_4, each with a lower and upper limit, the bounds on the feasibility region can be described by eight points. Each property constraint is translated into two inequality expressions from Eq. (3), hence there will be 2*NP* (number of properties) inequality equations that constitute the main set. The *AUP* values for these sets of equations will be calculated in order to determine the *AUP* range of the feasibility region. In a four property system, there will be 8 inequality equations in the original set, from which eight subsets will be developed. Each subset will be made up of four equations and only one of the two inequalities describing each property will be used in each subset. The following combination from the original set should be used to generate the eight subsets of equations:

$$\begin{matrix}(\Omega_1^{\max},\Omega_2^{\min},\Omega_3^{\min},\Omega_4^{\min}) \\ (\Omega_1^{\min},\Omega_2^{\max},\Omega_3^{\min},\Omega_4^{\min}) \\ (\Omega_1^{\min},\Omega_2^{\min},\Omega_3^{\max},\Omega_4^{\min}) \\ (\Omega_1^{\min},\Omega_2^{\min},\Omega_3^{\min},\Omega_4^{\max})\end{matrix} \quad , \quad \begin{matrix}(\Omega_1^{\min},\Omega_2^{\max},\Omega_3^{\max},\Omega_4^{\max}) \\ (\Omega_1^{\max},\Omega_2^{\min},\Omega_3^{\max},\Omega_4^{\max}) \\ (\Omega_1^{\max},\Omega_2^{\max},\Omega_3^{\min},\Omega_4^{\max}) \\ (\Omega_1^{\max},\Omega_2^{\max},\Omega_3^{\max},\Omega_4^{\min})\end{matrix} \tag{5}$$

The generated subsets of equations constitute the property constraints. In addition, structural constraints such as non-negativity constraints for the contribution of each group and a limit on the size of a molecular formulation need to be included. Finally, each group *g* has a free bond number (FBN) associated with it (e.g. CH_3 has FBN = 1, CH_2 has FBN=2), therefore a FBN constraint is needed to ensure that each designed molecule is structurally sound (see Eq. (6)). The general algebraic clustering procedure is given below:

$$n_g \geq 0 \quad , \quad \sum_{g=1}^{N_g} n_g \leq NF \quad , \quad FBN = \left[\sum_{g=1}^{N_g} n_g \cdot FBN_g\right] - 2 \cdot \left[\sum_{g=1}^{N_g} n_g - 1\right] \tag{6}$$

1. Transform given property data into molecular property operator terms.
2. Express property constraints as inequalities according to Eq. (4) to form the main set of equations. Determine the *AUP* range of the feasibility region.
3. Develop the subsets of inequality equations.
4. Generate the structural constraints according to Eq. (6).
5. Find the solution to each subset of linear inequality equations along with the structural constraint equations in order to determine the min and max n_g of each group *g*. This is done by first minimizing the *AUP* of each subset and then maximizing the *AUP* of each subset.

6. If the *AUP* values of each subset does not fall within the *AUP* range of the feasibility region, those solutions are excluded.
7. Solutions for n_g will not always be integer values, thus the solutions are rounded up for minimum values and rounded down for maximum values.
8. Generate all feasible formulations dictated by the solution and perform a final check that all property constraints are satisfied.

3. Application Example

To illustrate this new algebraic approach for molecular property clustering, a simple molecular synthesis problem is presented. Given CH_2 and OH groups as building blocks, the objective is to identify molecular formulations that will satisfy the following performance requirements:

$$\begin{aligned} 310 \le V_c(cm^3/mol) \le 610 \qquad & 90 \le H_v(kJ/mol) \le 120 \\ 20 \le H_{fus}(kJ/mol) \le 64 \qquad & 450 \le T_b(K) \le 560 \end{aligned} \tag{7}$$

Group contribution property data for the molecular groups is given in Table 1 [5,6]. The additive rules for the molecular operators are listed in Eq. (8) and (9).

$$V_c - v_{c0} = \sum_{g=1}^{N_g} n_g \cdot v_{c1} \quad , \quad H_v - h_{v0} = \sum_{g=1}^{N_g} n_g \cdot h_{v1} \tag{8}$$

$$H_{fus} - h_{fus0} = \sum_{g=1}^{N_g} n_g \cdot h_{fus1} \quad , \quad \exp\left(\frac{T_b}{t_{bo}}\right) = \sum_{g=1}^{N_g} n_g \cdot t_{b1} \tag{9}$$

Table 1. Property data for each molecular fragment

g	Group	FBN	V_c (cm^3/mol)	H_v (kJ/mol)	H_{fus} (kJ/mol)	T_b (K)
1	CH_2	2	56.28	4.91	2.64	0.9225
2	OH	1	30.61	24.21	4.79	3.21

Given the information in Table 1 and Eq. (8) and (9), the data for the four properties: critical volume, heat of vaporization, heat of fusion and boiling point are translated using the normalized property operator definition in Eq. (3) using with the following reference values (20, 1.0, 0.5, 7.0), respectively. The same reference values are also used to convert the group data given in Table 1. The resulting Ω values for all four property constraints are shown in Table 2. The *AUP* range of the feasibility region was calculated to be between 141.1 – 273.3.

Table 2. Dimensionless property operator values

	Ω_{Vc}	Ω_{Hv}	Ω_{Hfus}	Ω_{Tb}
Ω^{min}	15.105	78.26	45.612	1.291
Ω^{max}	30.102	108.26	4.133	2.213

Next, Eq. (4) is used to generate the main set of linear inequality equations, from which eight subsets are generated. The equations included in subset one according to Eq. (5) are provided below in Eq. (10). Finally the structural constraints are given in Eq. (11) and (12).

$$\begin{aligned} &1.53 \cdot g_1 + 2.81 \cdot g_2 \le 30.10 \\ &24.21 \cdot g_1 + 4.91 \cdot g_2 \ge 78.26 \\ &9.57 \cdot g_1 + 5.28 \cdot g_2 \ge 45.61 \\ &0.459 \cdot g_1 + 0.131 \cdot g_2 \ge 1.29 \end{aligned} \tag{10}$$

$$\begin{aligned} &g_1 \ge 0 \\ &g_2 \ge 0 \\ &g_1 + g_2 \le 15 \end{aligned} \tag{11}$$

$$\left[g_1 \cdot FBN_1 + g_2 \cdot FBN_2\right] - 2 \cdot \left[g_1 + g_2 - 1\right] = 0 \tag{12}$$

The results from solving the subsets of equations are summarized in Table 3. The solutions to the minimization problem of subsets 2, 5, 7 and 8 are excluded because their *AUP* values are outside the *AUP* range of the feasibility region. The results show that HO-$(CH_2)_7$-OH, HO-$(CH_2)_8$-OH, and HO-$(CH_2)_9$-OH are the formulations that satisfy all of the property and structural constraints.

4. Conclusions

In this work, an algebraic technique for molecular design within the property clustering framework has been introduced. Using the recently developed concepts of molecular property operators [4], this algebraic approach extends the application range of the original methodology to include more than three properties. The design problem is formulated as a set of linear algebraic equations analogous to the algebraic clustering technique developed for process design problems [2]. A simple proof of concept example described by four properties was solved successfully using this technique. The developed algebraic approach can be applied to problems that require both a higher

number of properties as well as additional groups. The resulting optimization problems are simply larger, but would still consist of linear algebraic equations.

Table 3. Results from solving the algebraic molecular synthesis problem

Subset	g_1+g_2	g_1	g_2	Objective	FBN	Ω_1	Ω_2	Ω_3	Ω_4	AUP
1	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
~~2~~	~~7~~	~~5~~	~~2~~	~~min~~	~~0~~	~~17.2~~	~~73.1~~	~~45.6~~	~~1.6~~	~~137.4~~
	14.2	12.2	2	max	0	37.4	108.3	83.5	2.5	231.6
3	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	15	13	2	max	0	39.6	112.3	87.8	2.6	242.3
4	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	15	13	2	max	0	39.6	112.3	87.8	2.6	242.3
~~5~~	~~6.3~~	~~4.3~~	~~2~~	~~min~~	~~0~~	~~15.1~~	~~69.4~~	~~41.7~~	~~1.5~~	~~127.8~~
	11.8	9.8	2	max	0	30.7	96.7	71.0	2.2	200.6
6	8.1	6.1	2	min	0	20.2	78.3	51.2	1.7	151.4
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
~~7~~	~~7~~	~~5~~	~~2~~	~~min~~	~~0~~	~~17.2~~	~~73.1~~	~~45.6~~	~~1.6~~	~~137.4~~
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8
8	~~4.8~~	~~2.8~~	~~2~~	~~min~~	~~0~~	~~11.0~~	~~62.3~~	~~34.1~~	~~1.3~~	~~108.8~~
	11.6	9.6	2	max	0	30.1	95.6	69.9	2.2	197.8

Acknowledgements

Financial support for this work has been provided by the National Science Foundation (NSF) CAREER program (CTS-0546925).

References

1. M. R. Eden, S. B. Jørgensen, R. Gani and M. M. El-Halwagi, *Chemical Engineering and Processing*, 43, 2004.
2. X. Qin, F. Gabriel, D. Harell, and M. M. El-Halwagi, *Industrial and Engineering Chemistry Research*, 43, 2004.
3. M. M. El-Halwagi, I. M. Glasgow, M. R. Eden and X. Qin, *AIChE Journal*, 50(8), 2004.
4. F. T. Eljack, M. R. Eden, V. Kazantzi, and M. M. El-Halwagi, *Computer Aided Chemical Engineering* 21A, 2006.
5. L. Constantinou and R. Gani, *AIChE Journal*, 40, 1994.
6. J. Marrero and R. Gani, *Fluid Phase Equilibria*, 182-183, 2001.