DECOMPOSED CAMD STRATEGY FOR SOLVENT MIXTURE DESIGN

Arunprakash T. Karunanithi and Luke E. K. Achenie
Department of chemical engineering, University of Connecticut
Storrs, CT-06269

Rafiqul Gani
CAPEC, Department of chemical engineering, Technical university of Denmark
Lyngby, Denmark

Abstract

Solvents and solvent formulations are widely used in many processes and products. Computer aided molecular/mixture design (CAMD) is one of the most common techniques used for the design of solvents and mixtures and/or formulations for different processes. Usually the CAMD problem is formulated as a mathematical programming problem (MINLP) and solved. In this paper the CAMD problem formulated as an MINLP is solved using decomposition based strategy. An industrial case study involving the design of solvent formulation for a pharmaceutical compound is presented.

Keywords

Molecular design, Mathematical programming, MINLP, Optimization.

Introduction

Solvents are widely used in industries for variety of purposes. Some of the important applications are extraction, crystallization, absorption etc. In many cases solvent mixtures are used instead of single solvents. Solvents are used not only for processing of products but in some cases as part of the final product formulation. Some examples of formulated products are pesticide formulation, paint formulations and drug formulations. Search for new solvents is one of the most important activities. This is driven by two things 1) the need to replace environmentally harmful solvents with environmentally benign ones 2) The need to identify solvents for processing new products. This need is prevalent particularly in pharmaceutical and bioprocess industries where new drugs are discovered every day and the level of toxic substances is bare minimum. Usually solvent selection is carried out by searching a solvent database. In this approach one cannot always find solvents that match the property requirements. A more systematic way of solvent selection can be carried out through computer aided molecular design (CAMD) approach. CAMD involves designing new chemicals that meet specified requirements. CAMD has been used before for the design of solvents (Gani and Brignole, 1983; Brignole et al., 1986; Buxton et al., 1999; Sinha and Achenie, 2001). Achenie at al., (2002) have defined computer aided molecular design as 'Given a set of building blocks and a set of target properties, determine the molecule or molecular structure that matches these properties'. In this technique the reverse problem of property estimation is tackled, that is, for a specified set of properties (target properties), pure compounds or mixtures that satisfy the property requirements are determined. CAMD primarily refers to pure component design. Designing solvent mixture replacement if no single chemical replacement is found or available can be termed as computer aided mixture design. Here the key issues are identifying the pure components and their compositions in the mixture.

Computer aided mixture design can be defined as ' Given a set of chemicals and a specified set of property constraints determine the optimal mixture'. In our approach the pure components are designed by molecular design and then the optimal mixture is identified. Usually the mixture design problem is posed as a Mixed Integer Nonlinear Optimization Programming (MINLP) problem where an objective function is optimized subject to constraints. The objective function is a process or product performance index and the constraints are usually molecular generation rules, process models, and product characteristics. For practical solvent mixture design problems, solving the generic MINLP problem has some difficulties, for example, the need to use highly complex property models for the prediction of product characteristics, the need to relate the product function with the process models, the search space, global optimal issues etc. In this work a decomposition based CAMD methodology is used where, the general MINLP model is decomposed into an ordered set of sub-problems, without changing the problem definition and solved as a series of sub-problems.

Methodology

A general mixture design problem can be represented as below

Min/Max $f_{obj}(X,Y)$

S.t *Structural constraints*: $g_1(Y) \le 0$

Pure component property constraints: $g_2(Y) \le 0$

Mixture property constraints: $g_3(X,Y) \le 0$

Process model constraints: $g_4(X,Y) = 0$

Y is a vector of integer variables, which are related to the identities of the building blocks of the pure component molecules. X is a vector of continuous variables, which are related to the mixture (e.g., compositions) and/or process variables(e.g., flow rates, temperatures etc.). f_{obj} is the objective function, which defines the optimization objective in terms of mixture-process (performance) characteristics and/or cost that may be minimized or maximized. g_1 and g_2 are structural constraints(related to feasibility of molecular structure) and pure component property constraints (related to properties-molecular structure relationships) respectively, which are function of integer variables alone. g_3 and g_4 are mixture property constraints (related to properties-mixture relationships) and process model constraints (related to process-molecule/mixture relationships) respectively, which are function of both integer and continuous variables.

The mixture design problem is divided into two parts. In the first part promising pure component solvents are designed and in the second part the solvent mixture is identified. The mixture design problem is solved as a series of sub-problems. The first three sub-problems which deal with the design of pure component solvents constitute the first part while the final two sub-problems that deal with mixture design constitute the second part. The first sub-problem considers the structural constraints that result in generation of feasible molecular structures. The second sub-problem considers the pure component properties, and the feasible molecular structures from the first sub-problem are solved for the pure component properties. Those molecules, which satisfy the pure component property constraints, are then passed into the third sub-problem, which considers the mixture properties related to pure component solvents. At the end of three sub-problems we will have all promising pure component solvents. In the fourth sub-problem mixture property constraints concerning the miscibility of the solvents is considered. In the final sub-problem the process model constraints are considered along with the objective function and the optimal mixture is identified by solving a MINLP problem or a series of NLP problems .In some cases such as product formulation we do not have process model and we would have to formulate an optimization problem with constraints other than process models in the final sub-problem.

Case study

The problem is to design a solvent formulation for a pharmaceutical compound. The compound has very low solubility in water. The aim is to identify an organic solvent which when added to water in small proportion will increase the solubility of the compound. The designed solvent should be completely miscible with water. It should be in liquid state at operating conditions and it should have high solubility for the solute. In this problem we do not have process model constraints. First the molecular design problem is solved to identify the optimal organic solvent that needs to be added with water.

Structural constraints

Sub-problem 1 considers the structural constraints. The structural constraints are imposed for a) ensuring that the number of bonds attached to a group equals the valence of the group; b) limiting number of groups in the molecule c) ensuring that the only one group is present in a position. For more details about structural constraints refer Churi and Achenie, 1996. The three structural constraints are stated below.

$$\sum_i \sum_j u_{ij}(2-\upsilon_j) = 2$$

$$\sum_i \sum_j u_{ij} = N_{\max}$$

$$\sum_j u_{ij} = 1$$

A multi level algorithm for generation of molecular structures proposed by Harper et al. was used. The requirement was to generate alcohols, aldehydes, ketones, acids, phenols, esters and ethers.

Pure component property constraints

In sub-problem 2 the pure component property constraints are considered. The three pure component property constraints are the upper limit and lower limit for the boiling point, melting point and solubility parameter of the solvent. Limits on boiling point and melting point are considered in order to ensure that the solvent is in liquid state at operating conditions. Usually if the solvent and the solute have similar solubility parameter values then the solubility of the solute in the solvent is high. The solubility parameter of the solute was estimated as 18.02 $Mpa^{1/2}$. Group contribution models (Constantinou and Gani, 1994) for estimating properties of pure compounds are used. The three pure component property constraints are shown below.

$$T_m = 102.425 * \sum_i N_i T_{mi} + \sum_j M_j T_{mj} \geq 380$$

$$T_b = 204.359 * \sum_i N_i T_{bi} + \sum_j M_j T_{bj} \leq 270$$

$$17 \leq \delta_t \leq 19$$

Mixture property constraints

For this problem there are no mixture property constraints related to the pure component solvent design. Hence sub-problem 3 is not considered

Miscibility constraint

In sub-problem 4 the requirement of complete miscibility of the organic solvent with water is considered. The stability function method gives the necessary and sufficient conditions for the phase stability of a binary pair in terms of mole fraction and activity coefficients (Bernard et al)

$$\frac{1}{x_2} + \frac{\partial \ln \gamma_2}{\partial x_2} \geq 0$$

If the above condition is satisfied for the entire composition range the pair is completely miscible. The activity coefficient is calculated using UNIFAC method (Freedunslund et al). Here it is worth noting that water has a solubility parameter value of 47.8 $Mpa^{1/2}$while the designed solvent will have a solubility parameter value between 17 and 19 $Mpa^{1/2}$.Consequently very few of the solvent candidates from the sub-problem 2 will satisfy the miscibility constraint in the sub-problem 4.

Sub-problem 5

The final optimization problem to be solved in sub-problem 5 is shown below

Objective function

$$f_{obj} = \min(x_2 \tau)$$

Constraints

$$\ln x_i^{Sat} - \frac{\Delta_{fus} H}{T_m}\left(1 - \frac{T_m}{T}\right) + \ln \gamma_1^{Sat} = 0$$

$$\frac{\Delta_{fus} H}{T_m}\left(1 - \frac{T_m}{T}\right) - \ln \gamma_1^{Sat} \geq 0.05$$

$$x_2 \leq 0.2$$

$$x_1 + x_2 + x_3 = 1$$

The objective is to minimze the toxicity of the final formulation. In this work –log(LC50) is considered as a quantitative measure of toxicity of the organic solvent. –log(LC50) values are evaluated using a group contribution method (Martin and Young, 2001). The first constraint is the solid liquid equilibrium constraint; the second constraint places a lower limit on the solubility; the third constraint places a upper limit on the composition of the organic solvent in the final formulation and the fourth constraint is for mole fraction satisfaction.

Results

In sub-problem 1, 3498 feasible molecular structures were generated. In sub-problem 2 only 645 out of 3498 molecules satisfied the pure component property constraints. Sub-problem 3 was ignored. In sub-problem 4 the 645 compounds were verified for the water miscibility. Among them only 5 compounds satisfied the constraint. Five NLP problems representing the five compounds were solved in sub-problem 5. The optimal molecule designed was 1,2-dimethoxy ethane. The structure of the solvent 1,2-Dimethoxy ethane is shown in figure 1.The pure component properties of the designed solvent are shown in

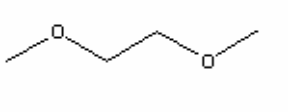

Figure 1

Table 1.Pure component properties

Property	Value	Unit
Solubility parameter at 298 K	17.93	$Mpa^{1/2}$
Melting Point	209.06	K
Boiling Point	377.82	K

Having identified the organic solvent as 1,2-dimethoxyethane the solubility of the drug at various compositions of water - 1, 2-dimethoxyethane mixtures was evaluated. These values are shown in table 2. The UNIFAC method was used for calculating the activity coefficient in determining the solubility.

Table 2.Solubility in water solvent mixtures

x_{sol}	x_{water}	Solubility of drug (mole percent)
0.05	0.95	1.59
0.10	0.90	3.48
0.15	0.85	4.64
0.20	0.80	5.34

The water solubility of the drug estimated using Marero Gani group contribution method was 0.0001 mole fraction .We can see from table 2 that addition of a small amount of the organic solvent improves the solubility to a great extent. Since the composition of the organic solvent in the final formulation needs to be very small because of toxicity reasons the solubility values were evaluated at lower compositions of the solvent.

Conclusions

A decomposition based methodology for solving computer aided molecular or mixture design problems is presented. It involves formulating the CAMD problem as an MINLP problem and then solving it as a series of sub-problems. These sub-problems are sub-set of constraints from the original set and are easy to solve. The problem becomes smaller and simpler as each sub-problem is solved. The industrial case study that was presented proved that this approach is flexible, effective and easy.

.

Nomenclature

T_b = Boiling point, K

T_m = Melting point, K

x_i = Molefraction of species i

δ_t = Solubility parameter, $Mpa^{1/2}$

γi = Activity coefficient of species i

References

Achenie, L. E. K., Gani, R. and Venkatsubramaniam, V., (2002). Computer aided molecular design: theory and practice, Elsevier.

Brignole, E. A., Bottini, S and Gani, R., (1986). A strategy for the design of solvents for separation processes, *Fluid Phase Equilibria*, 29, 125-132.

Buxton, A., Livingston, A.G. and Pistikopoulos, E.N., (1999). Optimal design of solvent blends for environmental impact minimization, *AIChE J.*, **45(4)**, 817-843.

Constantinou, L. and Gani, R., (1994). Group contribution methods for estimating properties of pure compounds. *AIChE J.*, **40(10)**, 1697-1710.

Frank, T. C., Downey, J. R. and Gupta, S.K. (1966). Quickly screen solvents for organic solids, *Chem.Eng.Prog.* ,**95 (12)**,41-61.

Fredenslund, A., Jones, R.L and Prausnitz, J.M., (1975). Group-Contribution estimation of activity coefficients in non ideal liquid mixtures, *AIChE J.*, **21(6),** 1086

Gani, R. and Brignole, E.A., (1983). Molecular design of solvents for liquid extraction based on UNIFAC, *Fluid phase equilibria*, 13, 331.

Gani, R. (2002). ICAS documentation, Internal report, CAPEC, Department of chemical engineering, Technical university of Denmark.

Gmehling, J.G., Anderson, T.F. and Prausnitz, J.M., (1975). Solid-liquid equilibria using UNIFAC, Ind.Eng.Chem.Fund., 17(4),269-273.

Sinha, M. and Achenie, L.E.K., 2001, Systematic design of blanket wash solvents with recovery considerations, Advances in environmental research, 5,239-249