Contents

I. Preface

1. Justification, objectives and contributions

  1.1. Justification
  1.2. Objectives
      1.2.1. Traditional and latent variable-based approaches applied to mixture design problems
      1.2.2. Latent variable-based approaches for efficient processes optimization
      1.2.3. Latent variable-based approaches applied to the Quality by Design initiative, to increase processes flexibility and guarantee the desired quality

  1.3. Contributions

2. On optimal design of experiments

  2.1. Introduction
  2.2. Optimality criteria for the design of experiments
      2.2.1. D-optimal design of experiments
      2.2.2. I-optimal and G-optimal design of experiments

  2.3. Additional considerations for the design of experiments
      2.3.1. Variance inflation
      2.3.2. Aliasing
      2.3.3. Principles of effect-sparsity, hierarchy and heredity

3. On latent variable and kernel-based multivariate data analysis

  3.1. Introduction
  3.2. Latent variable-based multivariate data analysis techniques
      3.2.1. Principal Component Analysis (PCA)
      3.2.2. Partial Least Squares regression (PLS)
  3.3. Kernel-based techniques
3.3.1. Basic principles of kernel-based techniques 21
3.3.2. Pseudo-samples and pseudo-sample projection 22
3.4. Important additional notions: cross-validation and jackknifing 23

4. Materials and methods 25
4.1. Hardware 25
4.2. Software 25
4.3. Datasets and methods 25

II. Mixture design optimization 27

5. Traditional approaches to mixture design 29
5.1. Introduction 29
5.2. The mixture space 31
   5.2.1. Assessing the shape of the mixture space 31
   5.2.2. Identifying the envelope of the mixture space 34
5.3. Regression model structures in mixture design 37
   5.3.1. The Scheffé models 37
   5.3.2. The Cox models 44
   5.3.3. Mixture-process variable models 46
5.4. Mixture design of experiments 48
   5.4.1. Simplex-based DOE 48
   5.4.2. DOE in irregular mixture spaces 53

6. Latent variable-based methods for mixture data analysis 57
6.1. Mixture design data analysis with Partial Least Squares 57
   6.1.1. Methods and datasets 58
      6.1.1.1. Example 1: seven-component octane blending experiment of Cornell 58
      6.1.1.2. Example 2: gasoline blending data of Snee 59
   6.1.2. Results and discussion 59
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4. DOE construction</td>
<td>100</td>
</tr>
<tr>
<td>7.5. Data analysis</td>
<td>103</td>
</tr>
<tr>
<td>7.5.1. Data analysis with MLR</td>
<td>110</td>
</tr>
<tr>
<td>7.5.2. Data analysis with PLS</td>
<td>112</td>
</tr>
<tr>
<td><strong>III. Design Space and optimization through the latent space</strong></td>
<td>113</td>
</tr>
<tr>
<td>8. Preliminary considerations</td>
<td>115</td>
</tr>
<tr>
<td>8.1. Quality by Design and the Design Space</td>
<td>115</td>
</tr>
<tr>
<td>8.2. Limitations of the optimization and DOE in the original space</td>
<td>118</td>
</tr>
<tr>
<td>8.3. Optimization in the latent space</td>
<td>119</td>
</tr>
<tr>
<td>9. Defining the design space in the latent space</td>
<td>123</td>
</tr>
<tr>
<td>9.1. Partial Least Squares model fitting and prediction uncertainty</td>
<td>123</td>
</tr>
<tr>
<td>9.2. Transferring restrictions to the latent space</td>
<td>125</td>
</tr>
<tr>
<td>9.3. Partial Least Squares model inversion</td>
<td>130</td>
</tr>
<tr>
<td>9.3.1. The direct inversion</td>
<td>130</td>
</tr>
<tr>
<td>9.3.2. Direct inversion-dependant definition of the Null Space</td>
<td>131</td>
</tr>
<tr>
<td>9.3.3. Analytical definition of the Null Space</td>
<td>132</td>
</tr>
<tr>
<td>9.3.4. Confidence region of the Null Space</td>
<td>134</td>
</tr>
<tr>
<td>9.4. Subspace most likely to contain the True Design Space</td>
<td>138</td>
</tr>
<tr>
<td>9.4.1. Datasets</td>
<td>139</td>
</tr>
<tr>
<td>9.4.1.1. Case study 1: mathematical model</td>
<td>139</td>
</tr>
<tr>
<td>9.4.1.2. Case study 2: simulated data with two correlated outputs</td>
<td>140</td>
</tr>
<tr>
<td>9.4.1.3. Case study 3: simulated Vinyl-Chloride Monomer manufacturing</td>
<td>141</td>
</tr>
<tr>
<td>9.4.2. Results</td>
<td>143</td>
</tr>
<tr>
<td>9.4.2.1. Case study 1: mathematical model</td>
<td>143</td>
</tr>
<tr>
<td>9.4.2.2. Case study 2: simulated data with two correlated outputs</td>
<td>145</td>
</tr>
</tbody>
</table>
9.4.2.3. Case study 3: simulated Vinyl-Chloride Monomer manufacturing

9.4.3. Conclusions

9.5. Subspace least likely to fall outside of the True Design Space

9.6. Assessing the adequacy of a PLS-regression model for inversion

9.6.1. Assessment via direct inversion

9.6.2. Assessment by comparison with the closest solution in the null-space

9.7. Additional considerations

10. Optimization problem formulation in Quality by Design

10.1. Introduction

10.1.1. Optimization in the original space through the latent space

10.1.2. Optimization in the latent space

10.2. Quadratic optimization formulation

10.2.1. Optimization of a linear combination of outputs

10.2.2. Optimization for exploration and DOE in the latent space

10.2.3. Tackling the maximization/minimization problem

10.2.3.1. Defining feasible minimum/maximum values as the desired ones

10.2.3.2. Changing the sign of the weight given in the objective function

10.2.3.3. Finding extreme achievable values below the Hotelling $T^2$ limit

10.3. Linear optimization formulation

10.4. A sequential optimization approach

11. Two real case studies of optimization in the latent space

11.1. Introduction

11.2. Methods

11.3. Datasets

11.3.1. Case study 1: minimizing two output variables simultaneously
11.3.2. Case study 2: maximizing a linear combination of outputs 178

11.4. Results and discussion 178
   11.4.1. Case study 1: minimizing two output variables simultaneously 178
   11.4.2. Case study 2: maximizing a linear combination of outputs 180

11.5. Conclusions 183

12. On experimentation to improve the design space estimation 185
   12.1. Introduction 185
   12.2. Methods 186
      12.2.1. Method 1: classical DOE and OLS model inversion approach 186
      12.2.2. Method 2: DOE in the latent space and LVRMI approach 189
      12.2.3. Method 3: Optimization in the latent space and LVRMI approach 191
   12.3. Datasets 192
      12.3.1. Case study 1: mathematical mod 192
      12.3.2. Case study 2: simulated Vinyl-Chloride Monomer manufacturing 192
   12.4. Results and discussion 193
      12.4.1. Case study 1: mathematical mod 193
         12.4.1.1. Detailed procedure for the application of all three methods 193
         12.4.1.2. Assessing the performance of each of the three methods 195
         12.4.1.3. Additional considerations 200
      12.4.2. Case study 2: simulated Vinyl-Chloride Monomer manufacturing 201
         12.4.2.1. Detailed procedure for the application of all three methods 201
         12.4.2.2. Assessing the performance of each of the three methods 203
   12.5. Conclusions 205
IV. Epilogue

13. Conclusions and perspectives

13.1. Accomplishment of the objectives

13.1.1. Objective I - Traditional and latent variable-based approaches applied to mixture design problems

13.1.2. Objective II - Latent variable-based approaches for efficient processes optimization

13.1.3. Objective III - Latent variable-based approaches applied to the Quality by Design initiative, to increase processes flexibility and guarantee the desired quality

13.2. Future research lines

14. Appendices

14.1. Annex to Part I

14.1.1. Relationship between the Euclidean distance matrix, \( D \), and the inner product matrix, \( X \cdot X^T \)

14.1.2. Practical meaning of the pseudo-samples in the feature space

14.2. Annex to Part II

14.2.1. Relationship between the Scheffé and Cox models coefficients

14.2.2. Projection of a point/vector onto the intersection of a group of hyperplanes

14.3. Annex to Part III

14.3.1. Relationship between the result of the PLS-regression direct inversion and the point in the combined null space closest to the centre of projection/with lowest leverage

14.3.2. Analytical expression for the confidence region of the null space for a linear combination of outputs using OLS-type expression for the prediction’s confidence interval

14.3.3. Confidence interval for the prediction of a linear combination of outputs in PLS using OLS type expression

Bibliography