Supplementary Material for "Physics Constrained Machine Learning for modeling of Chemical Refineries"

Rahul Golder[1], Bimol Nath Roy[1] and M. M. Faruque Hasan[1,2,*]
[1]Artie McFerrin Department of Chemical Engineering, Texas A&M University,
College Station, TX 77843-3122, USA.
[2]Texas A&M Energy Institute, Texas A&M University,
College Station, TX, 77843, USA.

1 Refinery Optimization Problem

This appendix provides the complete mathematical formulation, parameter sampling procedure, normalization, and feasibility restoration operator used for the refinery optimization problem discussed in Section "Refinery Optimization Problem".

1.1 Problem description

We consider a small refinery pooling problem with three feed streams, one intermediate pool, and two final gasoline products. The feed streams are straight-run naphtha (SRN), fluid catalytic cracked gasoline (FCC), and alkylate (ALK). The final products are regular gasoline and premium gasoline. The refinery decides how much of each feed stream to send to the pool, how much material to send from the pool to each product, and how much ALK to bypass directly to premium gasoline. The feed quality values are listed in Table 1, the product specifications are listed in Table 2, and the decision variables are listed in Table 3.

Table 1: Feed quality values for the refinery pooling problem.

Feed	Octane	Sulfur
SRN	72	25
FCC	92	100
ALK	94	5

Table 2: Product demand and quality specifications.

Product	Demand	Oct. min	Sul. max
Regular	$[80, 120]$	87	50
Premium	$[40, 80]$	91	30

*Correspondence should be addressed to M.M. Faruque Hasan at hasan@tamu.edu, Phone: (979) 862-1449.

Table 3: Decision variables for the refinery pooling problem.

Variable	Description
x_{SRN}	SRN flow to pool
x_{FCC}	FCC flow to pool
x_{ALK}	ALK flow to pool
y_{R}	Pool flow to regular gasoline
y_{P}	Pool flow to premium gasoline
z	ALK bypass flow to premium gasoline
p_O	Pool octane value
p_S	Pool sulfur value

1.2 Physical-space formulation

The objective is to maximize profit, defined as product revenue minus feed cost:

$$\begin{aligned}\Phi =& r_{\mathrm{R}} y_{\mathrm{R}} + r_{\mathrm{P}} \left(y_{\mathrm{P}} + z \right) \\ & - c_{\mathrm{SRN}} x_{\mathrm{SRN}} - c_{\mathrm{FCC}} x_{\mathrm{FCC}} - c_{\mathrm{ALK}} \left(x_{\mathrm{ALK}} + z \right).\end{aligned} \tag{1}$$

The feed availability constraints are

$$\begin{aligned} 0 \leq x_{\mathrm{SRN}} \leq A^U_{\mathrm{SRN}}, \\ 0 \leq x_{\mathrm{FCC}} \leq A^U_{\mathrm{FCC}}, \\ 0 \leq x_{\mathrm{ALK}} + z \leq A^U_{\mathrm{ALK}}. \end{aligned} \tag{2}$$

The pool capacity and product demand constraints are

$$\begin{aligned} x_{\mathrm{SRN}} + x_{\mathrm{FCC}} + x_{\mathrm{ALK}} \leq S, \\ 80 \leq y_{\mathrm{R}} \leq 120, \\ 40 \leq y_{\mathrm{P}} + z \leq 80. \end{aligned} \tag{3}$$

The material balance around the pool is

$$x_{\mathrm{SRN}} + x_{\mathrm{FCC}} + x_{\mathrm{ALK}} = y_{\mathrm{R}} + y_{\mathrm{P}}. \tag{4}$$

The pool octane and sulfur balances are

$$\begin{aligned} 72 x_{\mathrm{SRN}} + 92 x_{\mathrm{FCC}} + 94 x_{\mathrm{ALK}} = p_O \left(y_{\mathrm{R}} + y_{\mathrm{P}} \right), \\ 25 x_{\mathrm{SRN}} + 100 x_{\mathrm{FCC}} + 5 x_{\mathrm{ALK}} = p_S \left(y_{\mathrm{R}} + y_{\mathrm{P}} \right). \end{aligned} \tag{5}$$

The product quality constraints are

$$\begin{aligned} p_O y_{\mathrm{R}} \geq 87 y_{\mathrm{R}}, \\ p_S y_{\mathrm{R}} \leq 50 y_{\mathrm{R}}, \\ p_O y_{\mathrm{P}} + 94 z \geq 91 \left(y_{\mathrm{P}} + z \right), \\ p_S y_{\mathrm{P}} + 5 z \leq 30 \left(y_{\mathrm{P}} + z \right). \end{aligned} \tag{6}$$

Thus, the physical-space refinery optimization problem is

$$
\begin{aligned}
\max \quad & \Phi \\
\text{s.t.} \quad & (2) - (6).
\end{aligned}
\tag{7}
$$

The problem is nonlinear and nonconvex because the pool qualities p_O and p_S are decision variables that appear multiplied by product flow rates.

1.3 Parameter variation and data generation

For data generation, feed qualities and product specifications are fixed, while economic and availability parameters are varied. The sampled parameter vector is

$$
\theta = \begin{bmatrix} c_{\text{SRN}} & c_{\text{FCC}} & c_{\text{ALK}} & r_{\text{R}} & r_{\text{P}} & A^U_{\text{SRN}} & A^U_{\text{FCC}} & A^U_{\text{ALK}} & S \end{bmatrix}^T. \tag{8}
$$

The sampling bounds are listed in Table 4.

Table 4: Sampling bounds for varied refinery parameters.

Parameter	Lower	Upper
c_{SRN}	45	75
c_{FCC}	55	90
c_{ALK}	75	115
r_{R}	80	120
r_{P}	95	145
A^U_{SRN}	15	150
A^U_{FCC}	35	120
A^U_{ALK}	70	120
S	110	200

To avoid physically or economically meaningless instances, sampled parameters are required to satisfy

$$
\begin{aligned}
c_{\text{FCC}} &\geq c_{\text{SRN}} + 2, \\
c_{\text{ALK}} &\geq c_{\text{FCC}} + 2, \\
r_{\text{P}} &\geq r_{\text{R}} + 5, \\
r_{\text{R}} &\geq c_{\text{FCC}} + 3, \\
r_{\text{P}} &\geq c_{\text{ALK}} + 3, \\
A^U_{\text{SRN}} + A^U_{\text{FCC}} + A^U_{\text{ALK}} &\geq 120, \\
S + A^U_{\text{ALK}} &\geq 120.
\end{aligned}
\tag{9}
$$

Each sampled instance satisfying these rules is solved to generate training and testing data.

1.4 Normalized formulation

For a scalar quantity q with lower and upper bounds q^L and q^U, the normalized variable is defined as

$$
\hat{q} = \frac{q - q^L}{q^U - q^L}, \qquad q = q^L + \left(q^U - q^L\right)\hat{q}. \tag{10}
$$

The normalized sampled parameters are

$$
\begin{aligned}
c_{\mathrm{SRN}} &= 45 + 30\hat{c}_{\mathrm{SRN}},\\
c_{\mathrm{FCC}} &= 55 + 35\hat{c}_{\mathrm{FCC}},\\
c_{\mathrm{ALK}} &= 75 + 40\hat{c}_{\mathrm{ALK}},\\
r_{\mathrm{R}} &= 80 + 40\hat{r}_{\mathrm{R}},\\
r_{\mathrm{P}} &= 95 + 50\hat{r}_{\mathrm{P}},\\
A^{U}_{\mathrm{SRN}} &= 15 + 135\hat{A}_{\mathrm{SRN}},\\
A^{U}_{\mathrm{FCC}} &= 35 + 85\hat{A}_{\mathrm{FCC}},\\
A^{U}_{\mathrm{ALK}} &= 70 + 50\hat{A}_{\mathrm{ALK}},\\
S &= 110 + 90\hat{S}.
\end{aligned}
\tag{11}
$$

The normalized decision variables are

$$
\begin{aligned}
x_{\mathrm{SRN}} &= 150\hat{x}_{\mathrm{SRN}},\\
x_{\mathrm{FCC}} &= 120\hat{x}_{\mathrm{FCC}},\\
x_{\mathrm{ALK}} &= 120\hat{x}_{\mathrm{ALK}},\\
y_{\mathrm{R}} &= 120\hat{y}_{\mathrm{R}},\\
y_{\mathrm{P}} &= 80\hat{y}_{\mathrm{P}},\\
z &= 80\hat{z},\\
p_{O} &= 72 + 22\hat{p}_{O},\\
p_{S} &= 5 + 95\hat{p}_{S}.
\end{aligned}
\tag{12}
$$

The normalized decision vector is

$$
\hat{\boldsymbol{y}} = \begin{bmatrix} \hat{x}_{\mathrm{SRN}} \\ \hat{x}_{\mathrm{FCC}} \\ \hat{x}_{\mathrm{ALK}} \\ \hat{y}_{\mathrm{R}} \\ \hat{y}_{\mathrm{P}} \\ \hat{z} \\ \hat{p}_{O} \\ \hat{p}_{S} \end{bmatrix}, \qquad 0 \le \hat{\boldsymbol{y}} \le 1. \tag{13}
$$

For compactness, define

$$
Q = 120\hat{y}_{\mathrm{R}} + 80\hat{y}_{\mathrm{P}}. \tag{14}
$$

The normalized feed availability constraints are

$$
\begin{aligned}
150\hat{x}_{\mathrm{SRN}} &\le 15 + 135\hat{A}_{\mathrm{SRN}},\\
120\hat{x}_{\mathrm{FCC}} &\le 35 + 85\hat{A}_{\mathrm{FCC}},\\
120\hat{x}_{\mathrm{ALK}} + 80\hat{z} &\le 70 + 50\hat{A}_{\mathrm{ALK}}.
\end{aligned}
\tag{15}
$$

The normalized capacity and demand constraints are

$$
\begin{aligned}
150\hat{x}_{\mathrm{SRN}} + 120\hat{x}_{\mathrm{FCC}} + 120\hat{x}_{\mathrm{ALK}} &\le 110 + 90\hat{S},\\
80 \le 120\hat{y}_{\mathrm{R}} &\le 120,\\
40 \le 80\hat{y}_{\mathrm{P}} + 80\hat{z} &\le 80.
\end{aligned}
\tag{16}
$$

The normalized material balance is

$$150\hat{x}_{\mathrm{SRN}} + 120\hat{x}_{\mathrm{FCC}} + 120\hat{x}_{\mathrm{ALK}} = Q. \tag{17}$$

The normalized pool quality balances are

$$\begin{aligned} 10800\hat{x}_{\mathrm{SRN}} + 11040\hat{x}_{\mathrm{FCC}} + 11280\hat{x}_{\mathrm{ALK}} &= (72 + 22\hat{p}_O)\, Q, \\ 3750\hat{x}_{\mathrm{SRN}} + 12000\hat{x}_{\mathrm{FCC}} + 600\hat{x}_{\mathrm{ALK}} &= (5 + 95\hat{p}_S)\, Q. \end{aligned} \tag{18}$$

The normalized regular gasoline quality constraints are

$$\begin{aligned} (22\hat{p}_O - 15)\, \hat{y}_{\mathrm{R}} &\geq 0, \\ (95\hat{p}_S - 45)\, \hat{y}_{\mathrm{R}} &\leq 0. \end{aligned} \tag{19}$$

The normalized premium gasoline quality constraints are

$$\begin{aligned} (22\hat{p}_O - 19)\, \hat{y}_{\mathrm{P}} + 3\hat{z} &\geq 0, \\ (95\hat{p}_S - 25)\, \hat{y}_{\mathrm{P}} - 25\hat{z} &\leq 0. \end{aligned} \tag{20}$$