Parameter Estimation of Complicated Thermodynamic Models for Accurate Brine Separation

Pengfei Xu
Robert X. Gottlieb
PI: Prof. Matthew Stuber
Session: Industrial Applied Mathematics
Nov 7th, 2023
Brine separation is of critical importance to many industries with **brine effluent streams** and/or **brine concentration needs** (e.g., agriculture, power production, mining)

- Reduce costs
- Improve system robustness
- Increase sustainability

Motivation

Refined eNRTL

- Increase accuracy of thermodynamic properties calculation
- Improve accuracy of simulation results

\[ G^{*,\text{ex}} = G^{*,\text{SR}} + G^{*,\text{LR}} + \Delta G^{*,\text{Born}} \]

Short Range Interaction

Long Range Interaction

Born term, for aqueous system = 0

\[ = G^{*,\text{SR}} + (A^{*,\text{LR}} + PV) + 0 \]

Motivation

Modeling accuracy in high concentration regime

- Refined e-NRTL $^3$

Complexity of Refined eNRTL

\[
\frac{G_{SR}}{RT} = \sum_{j=1}^{n} X_{m_j} \left( \sum_{s\in\{m,a,c\}}^{n} \sum_{l=1}^{n} X_{s_l} F_{m_j, s_l, m_j, T} \right) \\
+ \sum_{j=1}^{n} X_{a_j} \sum_{k=1}^{n} X_{c_k} \left( \sum_{s\in\{m,c\}}^{n} \sum_{l=1}^{n} X_{s_l} F_{a_j, s_l, c_k} \right) \\
+ \sum_{j=1}^{n} X_{c_j} \sum_{k=1}^{n} X_{a_k} \left( \sum_{s\in\{m,a\}}^{n} \sum_{l=1}^{n} X_{s_l} F_{c_j, s_l, a_k} \right)
\]

\[
\log \gamma_{t_j}^{SR} = \frac{\partial}{\partial N_{t_j}} \left( \sum_{i\in\{m,a,c\}}^{n} \sum_{j=1}^{n} N_{i_j} \frac{G_{SR}}{RT} \right)
\]

Complexity of Refined eNRTL

\[ \frac{G^{SR}}{RT} = \sum_{j=1}^{n_m} X_{m_j} \left( \sum_{s \in \{m,a,c\}} \sum_{l=1}^{n_s} X_{s_l} F_{m_j,s_l,m_j} \right) \]

\[ + \sum_{j=1}^{n_a} \sum_{k=1}^{n_s} X_{e_k} \left( \sum_{s \in \{m,a,c\}} \sum_{l=1}^{n_s} X_{s_l} F_{a_j,s_l,a_k} \right) \]

\[ + \sum_{j=1}^{n_c} \sum_{k=1}^{n_s} X_{c_k} \left( \sum_{s \in \{m,a,c\}} \sum_{l=1}^{n_s} X_{s_l} F_{c_j,s_l,c_k} \right) \]

\[ \log \gamma_{ij}^{SR} = \frac{\partial}{\partial N_{t_j}^{i}} \left( \sum_{i \in \{m,a,c\}} \sum_{j=1}^{n_i} N_{t_j}^{i} \frac{G^{SR}}{RT} \right) \]

Automatic Differentiation Workflow

Symbolic Expression Generation

\[ \begin{align*}
G^{\text{SR}}, A^{\text{LR}}
\end{align*} \]

Automatic Differentiation (AD)

Symbolic Expression

\[ \log(\gamma_{t_j}^{\text{SR}}), \log(\gamma_{t_j}^{\text{LR}}) \rightarrow \log(\gamma_\pm) \]

Evaluate Expression

\[ \log(\gamma_\pm^{\text{Calc}}) \]

END

\( n_c, n_a, n_m \): Number of species in aqueous phase.

\( G^{\text{SR}} \): Short range excess Gibbs free energy.

\( A^{\text{LR}} \): Long range excess Helmholtz free energy.

\( \gamma_{t_j}^{\text{SR}} \): Short Range activity coefficient of species \( t_j \).

\( \gamma_{t_j}^{\text{LR}} \): Long Range activity coefficient of species \( t_j \).

\( t_j \): Species with type \( t \), \( t \in \{a, c, m\} \) and index \( j \).

$p^* \in \arg \min_{p \in P \subseteq \mathbb{R}^p} \sum_{i=1}^{n_d} \left( y_i(p) - y_i^{\text{data}} \right)^2$

s.t. $h(p) = 0$  \hspace{1cm} $g(p) \leq 0$

$n$, $n$, $n$: Number of species in aqueous phase.
$G^{\text{SR}}$: Short range excess Gibbs free energy.
$A^{\text{LR}}$: Long range excess Helmholtz free energy.
$\gamma^{\text{SR}}_t$: Short Range activity coefficient of species $t$.
$\gamma^{\text{LR}}_t$: Long Range activity coefficient of species $t$.
$t$: Species with type $t$, $t \in \{a, c, m\}$ and index $j$.
$N_i^{\text{Data}}$: Experimental data of concentrations of each species.
$\gamma^{\text{Data}}$: Experimental data of mean molal activity coefficient
$h$: Hydration number of each species, design variables for parameter estimation
$\tau$: Interaction parameters, design variables for parameter estimation
$N_i$: The concentration of each species in object system.
R-eNRTL Parameter Estimation

\[
\frac{G^\text{SR}}{RT} = \sum_{j=1}^{n_m} X_j^m \left( \sum_{s \in \{m,a,c\}} \sum_{l=1}^{n_i} X_{s_l}^j F_{s_l,m_j}^TF_{m_j,s_l,m_j}^T \right) + \sum_{j=1}^{n_a} X_{a_j} \sum_{k=1}^{n_c} X_{c_k} \left( \sum_{s' \in \{m,c\}} \sum_{l=1}^{n_i} X_{s'_l}^j F_{s'_l,a_j,c_k}^T \right)
\]

\[
+ \sum_{j=1}^{n_a} X_{a_j} \sum_{k=1}^{n_c} X_{c_k} \left( \sum_{s' \in \{m,c\}} \sum_{l=1}^{n_i} X_{s'_l}^j F_{s'_l,a_j,c_k}^T \right)
\]

\[
\sum_{s \in \{m,a,c\}} \sum_{l=1}^{n_i} X_{s_l}^j F_{s_l,a_j,c_k}^T \frac{\gamma_{a_j,c_k}}{X_{s_l}^j} \exp \left( -0.2 \tau_{a_j,s_l,c_k} \right)
\]

\[
p^* \in \arg \min_{p} \sum_{i=1}^{n_d} (y_i(p) - y_i^{\text{data}})^2
\]

s.t. \( h(p) = 0 \)

\( g(p) \leq 0 \)

\( \gamma \): Activity coefficient of species \( t \). 

\( N \): Experimental data of concentration of each species. 

\( \tau \): Interaction parameters, design variables for parameter estimation. 

\( \gamma_{a_j,c_k} \): Coefficient of species \( t \). 

\( N_j \): The concentration of each species in object system.

\( n, n_a, n_c, n_m \): Number of species in aqueous phase.

\( \gamma \): Short-range activity coefficient of species \( t \). 

\( \tau \): Interaction parameters, design variables for parameter estimation.

\( \gamma_{a_j,c_k} \): Coefficient of species \( t \). 

\( N_j \): The concentration of each species in object system.
EAGO.jl

Deterministic global optimizer
• High performance
• Open-source and free for non-commercial use
• Extensible
• Interval Arithmetic & McCormick based relaxation library

https://www.github.com/PSORLab/EAGO.jl

Challenge: Dependency Problem

\[ f(x) = \frac{\exp(x)}{\exp(x)} = 1 \]

\[ F(X) = \frac{\exp(X)}{\exp(X)} = \begin{bmatrix} \frac{\exp(x^L)}{\exp(x^U)} & \frac{\exp(x^U)}{\exp(x^L)} \end{bmatrix} \]

\[ X = [-5, 5] \]

\[ \begin{bmatrix} \frac{\exp(-5)}{\exp(5)} & \frac{\exp(5)}{\exp(-5)} \end{bmatrix} \]
Challenge: Dependency Problem

\[ f(x) = \frac{\exp(x)}{\exp(x)} = 1 \]

\[ F(X) = \frac{\exp(X)}{\exp(X)} = \left[ \frac{\exp(x^L)}{\exp(x^U)}, \frac{\exp(x^U)}{\exp(x^L)} \right] \]

\[ X = [-5, 5] \]

$$ \Rightarrow \left[ 4.5 \times 10^{-5}, 2.2 \times 10^4 \right] $$
Challenge: Dependency Problem

Abstracted Form of Problematic Terms

\[
\sum_{j=1}^{n} \exp(x_j)x_j
\]

EX: Single Term

\[
X = [x^L, x^U]
\]

\[
\frac{\exp(X)}{\exp(X)} = \left[\frac{\exp(x^L)}{\exp(x^L)}, \frac{\exp(x^U)}{\exp(x^L)}\right]
\]

\[
\frac{X \exp(X)}{\exp(X)} = \min \left\{ \frac{x^L \exp(x^U)}{\exp(x^L)}, \frac{x^L \exp(x^L)}{\exp(x^L)} \right\}, \max \left\{ \frac{x^U \exp(x^U)}{\exp(x^L)}, \frac{x^U \exp(x^L)}{\exp(x^L)} \right\}
\]
Function Profile \((n=3)\)

\[\begin{align*}
  x_3 &= -5 \\
  x_3 &= 0 \\
  x_3 &= 5
\end{align*}\]
Function Profile Animation

\[ x_3 = -5 \Rightarrow x_3 = 5 \]
\[ x_3 = -5.0 \]
Interval Extension \((n=3)\)

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} \]

**Domain**

1, 2, 3

\[ X_1, X_2, X_3 = [-5, 5] \]

**Interval Arithmetic**

\[ F \] \[ X = \left[ 3 \frac{-5 \exp(5)}{\exp(-5)}, 3 \frac{5 \exp(5)}{\exp(-5)} \right] \]

\[ \approx [-3 \times 10^5, 3 \times 10^5] \]

5 orders of magnitude larger than the interval hull of the image set

\[ \hat{f}(X) = [-5, 5] \]
A Tight Interval Extension Rule

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} = \frac{x_1 \exp(x_1) + x_2 \exp(x_2) + \ldots + x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} = \frac{\sum_{i=1}^{n} x_i \exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \]
A Tight Interval Extension Rule

\[
f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} = \frac{x_1 \exp(x_1) + x_2 \exp(x_2) + \ldots + x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} = \frac{\sum_{i=1}^{n} x_i \exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}
\]

\[
\sum_{i=1}^{n} \frac{x_i \exp(x_i)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} \cdot \frac{\exp(-x_i)}{\exp(-x_i)} = \sum_{i=1}^{n} \frac{x_i}{\exp(x_1 - x_i) + \exp(x_2 - x_i) + \ldots + \exp(x_n - x_i)}
\]
A Tight Interval Extension Rule

\[
f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} = \frac{x_1 \exp(x_1) + x_2 \exp(x_2) + \ldots + x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} = \sum_{i=1}^{n} \frac{x_i \exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)}
\]

\[
\sum_{i=1}^{n} \frac{x_i \exp(x_i)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} \cdot \frac{\exp(-x_i)}{\exp(-x_i)} = \sum_{i=1}^{n} \frac{x_i}{\exp(x_1 - x_i) + \exp(x_2 - x_i) + \ldots + \exp(x_n - x_i)}
\]

\[
\sum_{i=1}^{n} \frac{x_i}{\exp(x_1 - x_i) + \exp(x_2 - x_i) + \ldots + \exp(x_n - x_i)}
\]

\[
\Rightarrow f(x) = \sum_{i=1}^{n} \frac{x_i}{1 + \sum_{j=1}^{n} \exp(x_j - x_i)}
\]
A Tight Interval Extension Rule

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} = \frac{x_1 \exp(x_1) + x_2 \exp(x_2) + ... + x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + ... + \exp(x_n)} = \sum_{i=1}^{n} \frac{x_i \exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \]

\[ \sum_{i=1}^{n} \frac{x_i \exp(x_i)}{\exp(x_1) + \exp(x_2) + ... + \exp(x_n)} \cdot \exp(-x_i) = \sum_{i=1}^{n} \frac{x_i \exp(x_1 - x_i) + \exp(x_2 - x_i) + ... + \exp(x_n - x_i)}{\exp(x_1) + \exp(x_2) + ... + \exp(x_n)} \]

\[ \Rightarrow f(x) = \sum_{i=1}^{n} \frac{x_i}{1 + \sum_{j=1}^{n} \exp(x_j - x_i)} \]

- After conversion, although terms become more complex, the exponential terms in the numerator have been eliminated.
- Only denominator contains exponential terms.
A Tight Interval Extension Rule

\[ X_1, X_2, X_3 = [-5, 5] \]

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} \]

\[ f(x) = \sum_{i=1}^{3} \frac{x_i}{1 + \sum_{j=1, j\neq i}^{n} \exp(x_j - x_i)} \]

\[ F \quad X = \begin{bmatrix} 3 \frac{-5 \exp(5)}{\exp(-5)} , 3 \frac{5 \exp(5)}{\exp(-5)} \end{bmatrix} \]

\[ \approx [-3 \times 10^5, 3 \times 10^5] \]

\[ F \quad X = \begin{bmatrix} 3 \frac{-5}{1 + 2 \exp(-5 - 5)} , 3 \frac{5}{1 + 2 \exp(-5 - 5)} \end{bmatrix} \]

\[ \approx [-15, 15] \]
A Tight Interval Extension Rule

\[ X_1, X_2, X_3 = [-5, 5] \]

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} \]

\[ f(x) = \sum_{i=1}^{3} \frac{x_i}{1 + \sum_{j=1, j \neq i}^{n} \exp(x_j - x_i)} \]

\[ F \cdot X = \begin{bmatrix} \frac{3 -5 \exp(5)}{(-5)}, \frac{3 \exp(5)}{(-5)} \end{bmatrix} \]

Interval shrunk by 4 orders of magnitude.
McCormick Relaxations
Parameter Estimation Result

Parameter Estimation of Aqueous NaCl Solution Accuracy Verification

```
using JuMP, EAGO, ForwardDiff
import JuMP.@variable as @variable
new_fun = functionalize_obj_NaCl_fast(expr)
objfun(t1,t2,hc11,ha11) = objfun_eval_single_y^NaCl_fast(new_fun,t1,t2,hc11,ha11)
factory = () -> EAGO.Optimizer(SubSolvers())
model = Model(optimizer_with_attributes(factory, "absolute_tolerance" => 1e-4, "time_register" => model,:objfun,4,objfun,autodiff=true)
lb = [0. -10. 0. 0.]
ub = [10. 0. 2. 2.]
@variable(model, lb[i]< x[i=1:4] <= ub[i] )
@NLobjective(model, Min, objfun(x[1], x[2], x[3], x[4]))
@NLconstraint(model, y[1], x[3]-x[4]>=0)
optimize!(model)

✓ 9m 45.1s
```
Absolute Tolerance Achieved
First Solution Found at Node 1029
LBD = 0.0
UBD = 5.977314675518958e-6
Solution is:
\[ X[1] = 7.834167504201094 \]
\[ X[2] = -3.907172657309239 \]
\[ X[3] = 1.5789380575987575 \]
\[ X[4] = 0.723255866122256 \]
Scalability

Interaction Parameter

Neutral Species Centered (3 Types of Species Around)

\[ 2n_m n_a n_c \]

Cation Centered (2 Types of Species Around)

\[ \frac{1}{2} n_c n_a (n_a - 1) \]

Anion Centered (2 Types of Species Around)

\[ \frac{1}{2} n_a n_c (n_c - 1) \]

Hydration numbers \( h_c, h_a \)

Total number of parameters

\[
\frac{1}{2} n_c n_a (n_c - 1) + \frac{1}{2} n_a n_c (n_a - 1) + 2n_c n_a n_m + n_c + n_a
\]

For a system with 10 unique species each of anions and cations, water as solvent, there are 1120 parameters to fit.

AIChE 2023
Conclusions

➢ Implemented refined eNRTL model for single electrolyte case and generated activity coefficients as symbolic expressions using AD.

➢ Reformulated problematic multivariate quotient term to significantly reduce overestimation of dependency problems.

➢ Demonstrated the new rule by solving the parameter estimation problem for aqueous NaCl using EAGO.

➢ Currently implementing the multi-electrolyte form and expanding the AD work for all other thermodynamic properties
Thanks!

Members of the Process Systems and Operations Research Laboratory at the University of Connecticut (https://psor.uconn.edu/)

Funding:
National Alliance for Water Innovation (NAWI), DOE-EERE, Task 6.7

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Alliance for Water Innovation or the U.S. government.
Questions

Process Systems and Operations Research Laboratory

https://www.psor.uconn.edu
Thermodynamic of Refined eNRTL

\[ G^{*,\text{ex}} = G^{*,\text{SR}} + G^{*,\text{LR}} + \Delta G^{*,\text{Born}} \]

\[ \ln \gamma_j^*(T, P, x) \equiv \frac{1}{RT} \left( \frac{\partial G^{*,\text{ex}}}{\partial N_j} \right)_{T, P, N_{k\neq j}} \]

\[ = \frac{1}{RT} \left( \left( \frac{\partial G^{*,\text{SR}}}{\partial N_j} \right)_{T, P, N_{k\neq j}} + \left( \frac{\partial A^{*,\text{LR}}}{\partial N_j} \right)_{T, V, N_{k\neq j}} + \left( \frac{\partial A^{*,\text{LR}}}{\partial V} \right)_{T, P, N_j} \left( \frac{\partial V}{\partial N_j} \right)_{T, P, N_{k\neq j}} \right) \]
A Novel, Tight Interval Extension Rule

\[ f(x) = \frac{x^T \exp(x)}{1^T \exp(x)} = \frac{x_1 \exp(x_1) + x_2 \exp(x_2) + \ldots + x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} \]

\[ = \frac{x_1 \exp(x_1)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} + \frac{x_2 \exp(x_2)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} + \ldots + \frac{x_n \exp(x_n)}{\exp(x_1) + \exp(x_2) + \ldots + \exp(x_n)} \]

\[ = \frac{x_1}{1 + \exp(x_2 - x_1) + \ldots + \exp(x_n - x_1)} + \frac{x_2}{\exp(x_1 - x_2) + 1 + \ldots + \exp(x_n - x_2)} + \ldots + \frac{x_n}{\exp(x_1 - x_n) + \exp(x_2 - x_n) + \ldots + 1} \]

\[ = \sum_{i=1}^{n} \frac{x_i}{1 + \sum_{j=1, j \neq i}^{n} \exp(x_j - x_i)} \]
Challenge: Dependency Problem

Abstracted Form of Problematic Terms

\[
\left( \frac{\sum_{j=1}^{n} \exp(x_j x_j)}{\sum_{j=1}^{n} \exp(x_j)} \right)
\]

EX: 1 Term

\[
x_j = [x_j^L, x_j^U]
\]

\[
\exp(x_j) \in [\exp(x_j^L), \exp(x_j^U)]
\]

\[
x_j \exp(x_j) \in \left[ \min x_j^L \exp(x_j^U), x_j^L \exp(x_j^L) \right], \max x_j^U \exp(x_j^U), x_j^U \exp(x_j^L)
\]

\[
\frac{x_j \exp(x_j)}{\exp(x_j)} \in \left[ \min \left\{ \frac{x_j^L \exp(x_j^U)}{\exp(x_j^L)}, \frac{x_j^L \exp(x_j^L)}{\exp(x_j^L)} \right\} \right], \max \left\{ \frac{x_j^U \exp(x_j^U)}{\exp(x_j^L)}, \frac{x_j^U \exp(x_j^L)}{\exp(x_j^L)} \right\}
\]