Merging of the ‘Spline-Pointwise’ and ‘Morse/Long-Range’ Potential Function Forms for Direct-Potential-Fit Data Analyses

Jason Tao,\textsuperscript{a} Robert J. Le Roy\textsuperscript{a} and Asen Pashov\textsuperscript{b}

\textsuperscript{a} Guelph-Waterloo Centre for Graduate Work in Chemistry and Biochemistry, University of Waterloo, Waterloo, Ontario, Canada
\textsuperscript{b} Department of Physics, Sofia University, Sofia, Bulgaria

Research supported by the Natural Sciences and Engineering Research Council of Canada.
Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the ‘extrapolation region’ outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).
Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the ‘extrapolation region’ outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).

What is the best way of doing this?

Ans. By using a compact, analytical potential energy function
Objectives of Spectroscopic Data Analysis

1. To provide an accurate, compact, and comprehensive representation of experimental data.

2. To be able to interpolate reliably for missing observations within the data range.

3. To be able to provide realistic predictions in the ‘extrapolation region’ outside the range of existing data.

4. To provide reliable estimates of physically interesting molecular properties (e.g., bond lengths, force constants, intensities).

What is the best way of doing this?
Ans. By using a compact, analytical potential energy function

How do we determine this potential energy function?
Ans. By performing ‘direct potential fits’
Direct Potential Fits

• Simulate level energies as eigenvalues of some parameterized analytic potential energy function $V(r; \{p_j\})$

• Partial derivatives of observables w.r.t. parameters $p_j$ required for fitting are generated readily using the Hellmann-Feynman theorem:

$$\frac{\partial E(v, J)}{\partial p_j} = \left\langle \psi_{v,J} \left| \frac{\partial V(r; \{p_i\})}{\partial p_j} \right| \psi_{v,J} \right\rangle$$

• Compare predicted transition energies with experiment, and optimize potential parameters via an iterative least-squares fit
Direct Potential Fits

- Simulate level energies as eigenvalues of some parameterized analytic potential energy function $V(r; \{p_j\})$

- Partial derivatives of observables w.r.t. parameters $p_j$ required for fitting are generated readily using the Hellmann-Feynman theorem:

$$\frac{\partial E(v, J)}{\partial p_j} = \left\langle \psi_{v,J} \left| \frac{\partial V(r; \{p_i\})}{\partial p_j} \right| \psi_{v,J} \right\rangle$$

- Compare predicted transition energies with experiment, and optimize potential parameters via an iterative least-squares fit

**Challenge** ... to develop analytic potential function forms

* flexible enough to fully represent extensive high-resolution data
* robust and ‘well behaved’ (no spurious extrapolation behaviour)
* compact and portable – defined by a ‘modest’ no. of parameters
* incorporating appropriate physical limiting behaviour

Two successful approaches: 1. a *spline-pointwise* potential  
2. a global analytic function
1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (ad hoc) radial distance $r_{\text{out}}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{\text{out}}$) until a smooth connection is achieved.
1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (ad hoc) radial distance $r_{out}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{out}$) until a smooth connection is achieved.

Advantages

• Smooth and very flexible function: readily able to fit to irregular potentials such as double-minimum or ‘shelf’ potentials
• Little interparameter correlation
1. The Spline Pointwise Potential (SPP)

* $V(r)$ is represented by a cubic spline through a set of specified points
* The energies of the points are the fitted parameters
* Attach a long-range function at a chosen (ad hoc) radial distance $r_{\text{out}}$
* Fits/adjusts long-range coefficients (and sometimes also $r_{\text{out}}$) until a smooth connection is achieved.

Advantages

- Smooth and very flexible function: readily able to fit to irregular potentials such as those with double minima or a ‘shelf’
- Little interparameter correlation

Disadvantages

- Discontinuous derivatives at attachment to the extrapolation regions
- Third derivatives discontinuous at all spline points. Higher-order derivatives do not exist.
- Requires a large number of parameters/spline points ($\gtrsim 50$), each specified to many significant digits, making it inconvenient to copy and use
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = \mathcal{D}_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) \cdot y_p^{eq}(r)} \right)^2 \]

- \( y_p^{eq}(r) \) is the radial variable

\[ y_p^{eq}(r) = \frac{r^p - r_p^e}{r^p + r_p^e} \]
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = \mathcal{D}_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) \cdot y_{eq}(r)} \right)^2 \]

- \( y_{eq}(r) \) is the radial variable
  \[ y_{eq}(r) = \frac{r^p - r_e^p}{r^p + r_e^p} \]

- \( u_{LR}(r) \) defines long-range behaviour
  \[ u_{LR}(r) = \sum_{i=1}^{last} \frac{C_{m_i}}{r^{m_i}} \]
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = \mathcal{D}_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) y_{eq}^p(r)} \right)^2 \]

- \( y_{eq}^p(r) \) is the radial variable
  \[ y_{eq}^p(r) = \frac{r^p - r_e^p}{r^p + r_e^p} \]

- \( u_{LR}(r) \) defines long-range behaviour
  \[ u_{LR}(r) = \sum_{i=1}^{\text{last}} C_{m_i} \frac{1}{r^{m_i}} \]

- \( \beta(r) \) is the exponent coefficient function defined as
  \[ \beta(r) = \beta_q^p(r) = \beta_\infty y_{p \text{ref}}(r) + \left[ 1 - y_{p \text{ref}}^q(r) \right] \sum_{i=0}^{N} \beta_i [y_{q \text{ref}}^r(r)]^i \]

where the coefficients \( \beta_i \) are the fitting parameters and

\[ \lim_{r \to \infty} \beta(r) = \beta_\infty \equiv \ln \left( \frac{2\mathcal{D}_e}{u_{LR}(r_e)} \right) \]

these definitions allows the long-range behaviour of the potential to be

\[ V(r) \approx \mathcal{D}_e - u_{LR}(r) \]
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = D_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) \cdot y_{eq}^p(r)} \right)^2 \]

- \( y_{eq}^p(r) \) is the radial variable
  \[ y_{eq}^p(r) = \frac{r^p - r_e^p}{r^p + r_e^p} \]

- \( y_{eq}^p(r) \) is the radial variable
  \[ y_{eq}^p(r) = \frac{r^p - r_{ref}^p}{r^p + r_{ref}^p} \]

- \( u_{LR}(r) \) defines long-range behaviour
  \[ u_{LR}(r) = \sum_{i=1}^{last} \frac{C_{m_i}}{r^{m_i}} \]

- \( \beta(r) \) is the exponent coefficient function defined as
  \[ \beta(r) = \beta_p^q(r) = \beta_\infty y_{eq}^p(r) + [1 - y_{eq}^p(r)] \sum_{i=0}^{N} \beta_i [y_{eq}^q(r)]^i \]

where the coefficients \( \beta_i \) are the fitting parameters and

\[ \lim_{r \to \infty} \beta(r) = \beta_\infty \equiv \ln \left( \frac{2D_e}{u_{LR}(r_e)} \right) \]

these definitions allows the long-range behaviour of the potential to be

\[ V(r) \approx D_e - u_{LR}(r) \]
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = D_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) \cdot y_{eq}(r)} \right)^2 \]

**Advantages**

- Incorporates physically meaningful quantities \((D_e, r_e, C_n)\) as fitting parameters in the algebraic form
- Function and all derivatives smooth everywhere
- Requires relatively few parameters to achieve a better fit to experimental data than other forms

**Disadvantages**
2. Global Analytic Morse/Long-Range (MLR) Function

\[ V(r) = \mathcal{D}_e \left( 1 - \frac{u_{LR}(r)}{u_{LR}(r_e)} e^{-\beta(r) \cdot y_p^{eq}(r)} \right)^2 \]

**Advantages**

- Incorporates physically meaningful quantities (\( \mathcal{D}_e, r_e, C_n \)) as fitting parameters in the algebraic form
- Function and all derivatives smooth everywhere
- Requires relatively few parameters to achieve a better fit to experimental data than other forms

**Disadvantages**

- High correlation among parameters
- Difficulty accounting for abrupt changes in shape

*Can we combine the advantages of both forms?*
The Spline Exponent-MLR (SE-MLR)

Same structure as the MLR, except that it is $\beta(y_p^{\text{ref}}(r))$ [rather than $V(r)$] which is defined as a spline function through a specified set of function values, and it can be written as:

$$\beta(r) = \sum_{k=1}^{n} S_k(r) \beta(r_k) = \sum_{k=1}^{n} S_k(r) \beta_k$$

- Combines the very high flexibility of an SPP-type potential with the seamless incorporation of theoretical long-range behaviour inherent in the MLR form.
The Spline Exponent-MLR (SE-MLR)

Same structure as the MLR, except that it is \( \beta(y_p^{\text{ref}}(r)) \) [rather than \( V(r) \)] which is defined as a spline function through a specified set of function values, and it can be written as:

\[
\beta(r) = \sum_{k=1}^{n} S_k(r) \beta(r_k) = \sum_{k=1}^{n} S_k(r) \beta_k
\]

- Combines the very high flexibility of an SPP-type potential with the seamless incorporation of theoretical long-range behaviour inherent in the MLR form.

How do we use the SE-MLR?

1. Choose parameters defining \( y_p^{\text{ref}}(r) \) \{\( p \) and \( r_{\text{ref}} \)\}
2. Place the spline points \{\( N, y_p^{\text{ref}}(r_k), \beta_k \)\}
3. Fit to the data to optimize the \( \beta_k \) values
4. Consider the quality of fit (dimensionless root-mean-square deviation, \( \bar{dd} \)) and check the resulting potential for unphysical behaviour
Applications

In order to test the abilities of the SE-MLR, consider the following systems:

1. \( \text{Ca}_2 \ X^1\Sigma^+_g \)
   - 3573 data, uncertainties 0.006-0.15 cm\(^{-1}\)
   - Data covers 99.97% of \( \mathcal{D}_e \) (~1100 cm\(^{-1}\))
   - Highest observed level \( (v = 38) \) bound by only ~0.3 cm\(^{-1}\)
   - MLR treatments fitted \( C_6 \) while holding other dispersion coefficients \( (C_8, C_{10}) \) fixed

2. \( \text{N}_2 \ X^1\Sigma^+_g \)
   - 1221 data, uncertainties 0.0015-0.015 cm\(^{-1}\)
   - Data covers only 47% of \( \mathcal{D}_e \)
   - Highest observed level \( (v = 20) \) bound by 37600 cm\(^{-1}\)
   - **Challenge** Very narrow data region (0.9 - 1.55 Å), far extrapolation
How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\text{ref}}(r) \{ p, r_{\text{ref}} \}$

2. Place initial points \{here, 2 points < $r_e$, 13 points $\geq r_e$, $\beta(y_p^{\text{ref}}(\infty)) = \beta_\infty$\}
How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\text{ref}}(r) \{ p, r_{\text{ref}} \}$

2. Place initial points \{here, 2 points $< r_e$, 13 points $\geq r_e$, $\beta(y_p^{\text{ref}}(\infty)) = \beta_\infty$\}
How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\text{ref}}(r) \{ p \text{ and } r_{\text{ref}} \} \quad \checkmark$

2. Place the spline points $\{ N, y_p^{\text{ref}}(r_k), \beta_k \} \quad \checkmark$

3. Fit to the data to optimize the $\beta_k$ values \quad \checkmark

4. Consider the quality of fit (dimensionless root-mean-square deviation, $\overline{dd}$) and check the resulting potential for unphysical behaviour \quad \checkmark

5. Change number of points and repeat

6. Compare models with different numbers of points
How do we use the SE-MLR?

1. Choose parameters defining $y_p^{\text{ref}}(r)$ \{\(p\) and \(r_{\text{ref}}\)\}

2. Place the spline points \\{\(N\), \(y_p^{\text{ref}}(r_k)\), \(\beta_k\)\}

3. Fit to the data to optimize the \(\beta_k\) values

4. Consider the quality of fit (dimensionless root-mean-square deviation, \(\overline{dd}\)) and check the resulting potential for unphysical behaviour

5. Change number of points and repeat

6. Compare models with different numbers of points

![Graph showing the quality of fit for different numbers of points. The graph includes a recommended model highlighted.](image)
<table>
<thead>
<tr>
<th>Spline pointwise potential (2003)</th>
<th>PE – MLR\textsubscript{5,3}</th>
<th>SE – MLR\textsubscript{5}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{D}_e$</td>
<td>1102.060</td>
<td>1102.081</td>
</tr>
<tr>
<td>$r_{\text{out}}$</td>
<td>9.44</td>
<td>4.27781</td>
</tr>
<tr>
<td>$C_6$</td>
<td>$1.0023 \times 10^7$</td>
<td>$1.046 \times 10^7$</td>
</tr>
<tr>
<td>$C_8$</td>
<td>$3.808 \times 10^8$</td>
<td>$3.0608 \times 10^8$</td>
</tr>
<tr>
<td>$C_{10}$</td>
<td>$5.06 \times 10^9$</td>
<td>$8.344 \times 10^9$</td>
</tr>
</tbody>
</table>
| \begin{array}{|c|c|c|}
| $r_{\text{ref}}$ & 5.55 & \beta_0 & -0.19937072 \\
| \beta_1 & -0.23219 & \beta_2 & -0.06091 \\
| \beta_3 & 0.1383 & \beta_4 & -0.1791 \\
| \beta_5 & 0.362 | \beta_6 & 0.249 \\
| \hline
| dd & 0.628 & \end{array} |
| \begin{array}{|c|c|c|}
| $y_5^{\beta,3}$ & \beta & \end{array} |
| \begin{array}{|c|c|}
| -1.000 & 0.0084239 \\
| -0.844 & -0.0034414 \\
| -0.688 & -0.0300237 \\
| -0.447 & -0.0839978 \\
| -0.206 & -0.1384960 \\
| 0.034 & -0.1865833 \\
| 0.276 & -0.2250030 \\
| 0.517 & -0.2481562 \\
| 0.758 & -0.2466246 \\
| 1.000 & -0.2134933 \\
| \hline
| dd & 0.628 & \end{array} |

\begin{array}{|c|}
| dd & 0.70 |
| \end{array}
Results

The SE-MLR achieves accuracy of the PE-MLR with requiring significantly fewer parameters than the SPP, but more than the PE-MLR.

<table>
<thead>
<tr>
<th>Ca$_2$(X$^1\Sigma_g^+$)</th>
<th>SPP</th>
<th>PE-MLR</th>
<th>SE-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td># fitted param.</td>
<td>55</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>$dd$</td>
<td>0.70</td>
<td>0.628</td>
<td>0.628</td>
</tr>
</tbody>
</table>

The fitted value of $C_6$ obtained using the SE-MLR shows less model dependence than when using the PE-MLR.
Results

The SE-MLR achieves accuracy of the PE-MLR with requiring significantly fewer parameters than the SPP, but more than the PE-MLR.

<table>
<thead>
<tr>
<th></th>
<th>SPP</th>
<th>PE-MLR</th>
<th>SE-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca₂(X₁Σ⁺)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># fitted param.</td>
<td>55</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>dd</td>
<td>0.70</td>
<td>0.628</td>
<td>0.628</td>
</tr>
</tbody>
</table>

The fitted value of C₆ obtained using the SE-MLR shows less model dependence than when using the PE-MLR.

<table>
<thead>
<tr>
<th></th>
<th>PE-MLR</th>
<th>SE-MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₂(X₁Σ⁺)</td>
<td></td>
<td></td>
</tr>
<tr>
<td># fitted param.</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>dd</td>
<td>1.416</td>
<td>1.404</td>
</tr>
</tbody>
</table>

SE-MLR can incorporate a sensible long extrapolation to the limit, but a conventional SPP cannot!
A challenging System for the SE-MLR form

Double-minimum potential - $Na_2(2^1\Sigma_u^+)$
Conclusion

• By having the exponent coefficient function (rather than the potential itself) be represented by a cubic spline, the number of points required to describe the potential is reduced dramatically.

• The SE-MLR successfully combines the flexibility of the spline-pointwise approach with a natural incorporation of the theoretically predicted inverse-power long-range behaviour.

• To obtain a given quality of fit for a conventional single-minimum potential, the SE-MLR requires more parameters than does a PE-MLR.

• However, preliminary results suggest that fits using an SE-MLR may provide a more reliable determination of long-range coefficients such as $C_6$.

Future Work

• Explore the SE-MLR’s utility in describing double-minimum potentials.

• Possible quantitative incorporation of the correct, very short range ‘united-atom-limit behaviour’ $V(r) = \frac{Z_1 Z_2 e^2}{4 \pi \epsilon_0 r}$ ?
Splitting the radial variable exponent

\[ \beta(r) = c_0 y_q^{\text{ref}}(r) + c_1 y_q^{\text{ref}}(r) + c_2 y_q^{\text{ref}}(r)^2 + c_3 y_q^{\text{ref}}(r)^3 \]

\[ y_q^{\text{ref}}(r) = \frac{1 - \left(\frac{r}{r_{\text{ref}}}\right)^q}{1 + \left(\frac{r}{r_{\text{ref}}}\right)^q} = \left(\frac{1}{1 + x^q}\right) (1 - x^q) \]

For \( q < p \)

\[ e^{-\beta(y_q^{\text{ref}}(r))y_p^{\text{eq}}(r)} = e^{\beta_{\infty}} \left(1 + \frac{2r_q^{\text{ref}}(c_1 + 2c_2 + 3c_3)}{r^q} + \frac{2r_p^{\text{eq}} C}{r^p} + \frac{2r_{2q}^{\text{ref}}(c_1 - 3c_3)}{r^{2q}}\right) \]

However, at \( r = \infty \), \( y_q^{\text{ref}}(r) = 1 \) and if

\[ \frac{d\beta(y_q^{\text{ref}})}{dy_q^{\text{ref}}} = c_1 + 2c_2 + 3c_3 = 0 \]

then at large distances

\[ V(r) \approx D_c \left(1 - u_{LR}(r) + \frac{2r_e^{\beta_{\infty}}}{r^p} + \ldots\right) \]
Improving Short-range behaviour

$$\lim_{r \to 0} V(r) = \frac{C_1}{r} = \frac{Z_1 Z_2 C_1^{pp}}{r} \approx \frac{Z_1 Z_2 C_1^{pp} \mathcal{D}_e u_{LR}(r) e^{2\beta(r=0)}}{u_{LR}(r_e)^2 r}$$

$$\beta(r = 0) = \ln \left( \frac{u_{LR} \sqrt{Z_1 Z_2 C_1^{pp} / \mathcal{D}_e u_{LR}(r_e)}}{\sum_{i=1}^{\text{last}} \left( \frac{b \rho}{m_i} \right)^{m_i-1/2}} \right)$$