
Article:
Uteva, E., Graham, R.S., Wilkinson, R.D. orcid.org/0000-0001-7729-7023 et al. (1 more author) (2017) Interpolation of intermolecular potentials using Gaussian processes. Journal of Chemical Physics, 147 (16). 161706. ISSN 0021-9606

https://doi.org/10.1063/1.4986489

ChemComm

Interpolation of intermolecular potentials using Gaussian processes.

Journal:	*ChemComm*
Manuscript ID	CC-COM-01-2016-000093
Article Type:	Communication
Date Submitted by the Author:	05-Jan-2016
Complete List of Authors:	Uteva, Elena; University of Nottingham, School of Chemistry Graham, Richard; University of Nottingham, School of Mathematical Sciences Wilkinson, Richard; University of Sheffield, School of Mathematics and Statistics Wheatley, Richard; The University of Nottingham, School of Chemistry

Interpolation of intermolecular potentials using Gaussian processes.
Elena Uteva, Richard S. Graham, Richard D. Wilkinson and Richard J. Wheatley.

We hope that this manuscript can be published in Chemical Communications. Intermolecular potentials are ubiquitous in Chemistry and associated scientific areas, and while calculating them has become more routine over the last 10-20 years (at least for small and medium-sized rigid molecules, at a limited number of geometries), interpolating or fitting the calculated data to produce a complete multidimensional potential energy surface is a much more difficult problem, consuming a great amount of researcher time, with no satisfactory solution as yet.

We think that the method described herein will be the benchmark for interpolating intermolecular potential data. In particular, the innovation of using inverse internuclear distances as coordinates makes a dramatic difference to the results (see especially figure 1, but very similar figures could have been produced for all the studied molecules, given enough space). The methods described can be learned from scratch by a student in a few days, and used to design and interpolate a complete intermolecular potential energy surface in a few hours, compared to months of time spent in the past on producing usually inferior fits.

In short, we believe that this work represents a clear step-change in an important area of Chemistry, and we are pleased to submit it for your consideration.

Yours sincerely, Richard Wheatley.

Interpolation of intermolecular potentials using Gaussian processes.

Elena Uteva [a], Richard S. Graham [b], Richard D. Wilkinson [c] and Richard J. Wheatley[a*].

Received Xth XXXXXX 20XX, Accepted Xth XXXXXX 20XX
First published on the web Xth XXXXXXX 200X
DOI: 10.1039/b000000x

Abstract

A general procedure is proposed to produce intermolecular potential energy surfaces efficiently from a relatively small number of training data. The procedure involves generation of geometrical configurations using a Latin hypercube design, with a maximin criterion based on internuclear distances. Gaussian processes are used to interpolate the data, using over-specified inverse molecular distances as covariates, greatly improving the interpolation. Symmetric covariance functions are specified so that the interpolation surface obeys all relevant symmetries, reducing prediction errors. Results are presented for two systems involving CO_2, a system with a deep energy minimum (HF – HF) and a system with 48 symmetries ($CH_4 – N_2$). In each case the approach predicts an independent test set, with RMS error values that are comparable with or better than the best literature fits.

1 Introduction

Computational chemistry has advanced to the stage where calculations of intermolecular potential energies can be performed accurately enough, for small molecules, to be useful in areas including chemistry, physics, atmospheric science, geology and biochemistry. However, the computational cost of evaluating the energy at a single point in coordinate space is significant (often minutes or hours of time), so it is necessary to fit or interpolate calculated energy data to produce a potential energy surface for any intermolecular geometry of interest.

The choice of fitting or interpolation method, and the amount of data that are needed, are both significant limiting factors in the generation of accurate potential energy surfaces. Examples of careful and elaborate fits of calculated data include the potential energy surface of $CO_2 – Ne$ [1], where a root mean square error (RMSE) of about 0.15 μE_h was quoted ($E_h \approx 2625.5$ kJ mol^{-1}); a RMSE of about 0.6 μE_h in the well region (energy $E < 0$) of $CO_2 – H_2$ [2]; and a maximum error of about 2% of the well depth in the well region of $CH_4 – N_2$ [3]. Fits with much larger errors are commonplace in the literature, even when RMSE scores are based on the fit to training data, rather than independent test data, a procedure which is prone to over-estimating predictive accuracy. Interpolations of intermolecular potential data are less common. Cubic splines are the most popular interpolation method, for example in work on $CO_2 – Ar$ [4]. In contrast, Gaussian process (GP) interpolation, of which cubic splines are a special case, has been little used[5,6], despite its promise in other applications. The few applications include solid-state potentials [7,8], and the difference between calculated intermolecular potentials of water [9], but not interpolating a complete intermolecular potential energy surface. The development of a general interpolation method, which produces reliable results based on relatively few calculated energies, would constitute a major advance in this research area. It is demonstrated here that with a carefully chosen set of training points and coordinate system, symmetric Gaussian process interpolation of intermolecular potentials can achieve high predictive accuracy.

2 Gaussian process modelling

The approach involves two sets of data. A set of training data (between 20-1000 points) is used to train the model, and a larger set of grid data is used to test the model's predictive performance. Both datasets are described below. No knowledge of the test data is used during training.

2.1 Intermolecular potential data

Data sets of the intermolecular interaction energy of the bimolecular complexes $CO_2 – Ne$, $CO_2 – H_2$, HF – HF and $CH_4 – N_2$ are calculated as a function of their configurational geometry. All molecules are approximated as linear rigid rotors in their vibrational ground state, with fixed bond lengths. Energy calculations are carried out in Molpro [10] using second-order Möller-Plesset perturbation theory (MP2) and augmented correlation-consistent triple-zeta (aug-cc-pVTZ) basis sets. Basis set superposition errors are corrected using the full counterpoise correction procedure.

[a] *School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.*
[b] *School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK.*
[c] *School of Mathematics and Statistics, University of Sheffield, Western Bank Sheffield, S10 2TN UK.*

Jacobi coordinates are used to describe the multi-dimensional potential energy hypersurfaces (see Table 1). In all cases r is the distance between the molecular centres. For CO_2-Ne, θ is the angle between r and the CO_2 axis. For CO_2-H_2, θ_1 is the angle between r and the CO_2 axis, θ_2 is the angle between r and the H_2 axis, and ϕ is the torsional angle of the H_2 axis. Analogous coordinates are used for HF–HF. For CH_4-N_2, the N_2 molecule is placed at a position relative to the C of CH_4 at position (r,θ,ϕ) in polar coordinates, and the N-N axis is rotated to orientation (α,β), also in polar coordinates. The C-O, H-H, H-F, C-H and N-N bond lengths are taken to be 1.1632 Å, 0.77 Å, 0.92 Å, 1.09 Å and 1.098 Å, respectively. An energy cutoff of $E_{\rm cut} = 0.005\ E_{\rm h}$ is imposed ($0.02\ E_{\rm h}$ for HF-HF due to its much larger well depth), and molecular configurations with intermolecular potentials that exceed this cutoff are excluded from the data sets. Configurations are also excluded if any interatomic distance is below 1.5 Å or if all interatomic distances are above 8.5 Å. Separations below this would also be excluded by the energy cutoff, but this criterion saves time that would be spent in calculating unhelpfully large energies, and beyond 8.5 Å it is more efficient to use an asymptotic expansion of the energy, as discussed later. Details of the test data used for model assessment are given in Table 1.

Table 1 Coordinates for the test (grid or LHC) data for each system.

System	Test Grid or Latin Hypercube			Test
	Coordinate	Range	Spacing	points
CO_2-Ne	r	1.5-10 Å	0.116 Å	1122
	$\cos\theta$	0-1	0.05	
CO_2-H_2	r	1.5-10 Å	0.5 Å	12844
	$\cos\theta_1$	0-1	0.111	
	$\cos\theta_2$	0-1	0.111	
	ϕ	0-180°	20°	
HF–HF	Latin hypercube			2158
CH_4-N_2	Latin hypercube			1182

2.2 Gaussian process training

Gaussian processes (GPs)[11] are used extensively in machine learning and statistics as regression models. They are 'non-parametric' models of functions, which generalise the Gaussian distribution. The prior specification of a GP consists of a mean function (often taken to be zero) and a covariance function $k(\mathbf{x},\mathbf{x}')$, which expresses the covariance between $f(\mathbf{x})$ and $f(\mathbf{x}')$, where f is the function being interpolated. Training data, consisting of observations of the value of f at various locations, are used to update the mean and covariance functions to give a posterior model which can be used to predict the function at any location.

Properties of the resulting GP model are inherited from the covariance function, for example, differentiability, continuity and stationarity. The intermolecular energy is a non-stationary function of distance, as it varies rapidly at small interatomic separations, but more gently at larger separation. Although it is possible to write down a non-stationary covariance function, in practice it can be challenging to specify a flexible form that captures the correct non-stationary behaviour. It is simpler to transform either the inputs or outputs to achieve approximate stationarity, which is addressed here by using the inverse interatomic distances as covariates in the GP. Thus the GP coordinates are $\mathbf{x} = (1/r_1, ..., 1/r_{N_D})$ where r_i is the interatomic distance, running over all pairs of nuclei belonging to different molecules. This results in an over-specified system, for example with $N_D = 6$ dimensions for CO_2-H_2. It is shown later that this change in variables leads to a dramatic improvement in performance.

The training data should ideally comprise approximately evenly spaced points in a single symmetry-distinct sub-region of $\mathbf{x}$ space, and respect the geometric constraint. The general strategy is to generate many candidate data sets (coordinates only, not energies), exclude points outside the symmetric and geometric constraints, and select the candidate data set with the best distribution of points. Specifically, for CO_2-Ne and CO_2-H_2, candidate data sets are generated from Latin hypercube (LHC) sampling of $1/r$ and the angular LHC coordinates in Table 1. For HF–HF, three LHCs are generated and combined into one dataset: one uses the F-F distance as the radial coordinate r, and keeps only those data points within the LHC for which the F-F distance is the shortest of the four internuclear distances; the other two LHCs are generated in the same way but with F-F replaced by H-H and H-F in turn. The LHC for CH_4-N_2 is generated based on an H-N distance as the radial coordinate, and uses only the data points for which the same H-N distance is the shortest of the ten internuclear distances. For all four interactions, after generating the LHCs, deleting data points based on the symmetric and geometric constraints, and combining the sets of points into one (for HF-HF), the minimum separation of the remaining points is calculated in $\mathbf{x}$ space. The candidate data set with the largest minimum separation is used as the training set. This 'maximin' approach aims to give even coverage across the whole of the relevant region of $\mathbf{x}$ space.

The Gaussian process has a zero mean function and a squared-exponential covariance function

$$\kappa(\mathbf{x},\mathbf{x}') = \sigma_f^2 \prod_{i=1}^{N_D} \exp\left[-\frac{(x_i - x_i')^2}{2l_i^2}\right] \qquad (1)$$

where σ_f^2 is the signal variance and l_i is the correlation length for each dimension. This choice results in a stationary infinitely differentiable model, which is called the 'non-symmetric model'.

The potential energy surfaces obey various symmetries in $\mathbf{x}$ space. For example, for CO_2-Ne, the energy is invariant under the interchange of the two coordinates corresponding to distances between Ne and each of the O atoms. Let G represent the permutation group containing permutations of elements of $\mathbf{x}$ under which the energy surface is unchanged. If it is assumed that $l_i = l_j$ when coordinates x_i and x_j swap for some permutation in G, then a covariance function of the form

$$k_{\mathrm{sym}}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \sum_{g\in G} \kappa(g\mathbf{x},\mathbf{x}'). \qquad (2)$$

results in a GP which shares the symmetries of the energy surface (see the Supplementary Material). The 'symmetric model' based on this symmetric covariance function gives predictions that respect the relevant symmetries, and usually significantly improves the performance, even within the symmetry-unique region covered by the test data, as shown below.

The GPs are obtained using the GPy package[12] modified to include symmetric covariance functions. Zero-mean Gaussian observation error[11] is assumed on the function outputs (refered to as *nugget* in geostatistics), with standard deviation σ_n. Thus the model's hyperparameters are σ_f, σ_n and $\{l_i\}$. These hyperparameters are estimated by optimising the log-likelihood over ≈ 30 random restarts, which typically is sufficient to find the optimal values multiple times.

The choice of inverse internuclear distances to transform to stationarity is important. To illustrate this a 'basic model' GP is created, which uses internuclear distances r as coordinates rather than $1/r$, but is otherwise identical to the non-symmetric GP above. In particular, the same test and training data are used, and the covariance function has the same form as equation (1).

3 Results

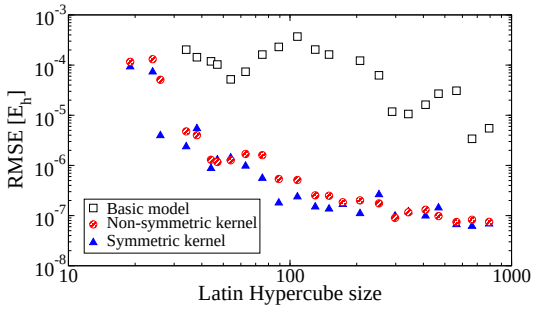

Fig. 1 RMSE against LHC size for CO_2-Ne. The lowest energy in the grid data is $-2.90\times10^{-4}\ E_h$.

Predictive performance is measured using the root mean square error (RMSE) of the GP predictions of the test data. Note that the GP has no advance knowledge of the test data, only the far more limited training data. The RMSEs are somewhat noisy because of the random nature of generating LHCs, and because relatively small fractional errors in individual points high on the repulsive wall have a significant effect on the RMSE. However, this variability is usually small compared to the effect of increasing the LHC size, as demonstrated next.

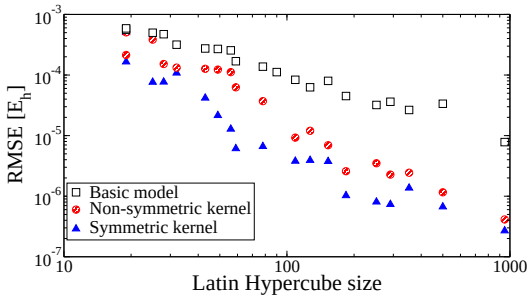

Fig. 2 RMSE against LHC size for CO_2-H_2. The lowest energy in the grid data is $-8.25\times10^{-4}\ E_h$.

The results for CO_2-Ne are shown in Figure 1. Here, the models based on inverse intermolecular distances dramatically outperform the basic model, being typically 2-3 orders of magnitude more accurate, when compared at fixed LHC size. Furthermore, even though this system contains only one symmetry, the symmetric model is typically a factor of 2 more accurate than the non-symmetric model. Figure 2 shows similar results for CO_2-H_2. Here, the inverse distance models again strongly outperform the basic model, achieving RMSEs $< 10^{-6}\ E_h$ for a reasonable number of training points. The symmetric kernel typically gives a factor of 2-10 improvement, with the greater improvement compared to CO_2-Ne probably resulting from the greater number of symmetries.

For $HF-HF$, the minimum energy in the calculated test data is $-6.17\times10^{-3}\ E_h$, which is about an order of magnitude larger than for the other interactions. Probably as a consequence of this, it is found to be necessary to include training points up to at least $10^{-2}\ E_h$, otherwise the prediction of the few remaining points on the repulsive wall is poor. Using a cutoff of $2\times10^{-2}\ E_h$ gives an RMSE of $1.6\times10^{-4}\ E_h$ for a symmetric GP with 59 training points, and the RMSE generally decreases with increasing numbers of training points, to $1.8\times10^{-5}\ E_h$ for 327 training points. The RMSE in the negative-energy region is about $5\times10^{-6}\ E_h$ for the latter GP; one or two high-energy points dominate the overall RMSE. The inclusion of symmetry in the GP has little effect on the RMSE for this interaction.

For CH_4-N_2, all 48 symmetry elements are included in the GP. The minimum energy in the test data is $-6.98\times10^{-4}\ E_h$. As might be expected, the inclusion of symmetry is impor-

tant for this interaction, even though all the training and test data are confined within a single symmetry-distinct region of space. With a training set of 106 points, the RMSE is found to be $51 \times 10^{-6}\ E_h$ for the nonsymmetric GP and $6.8 \times 10^{-6}\ E_h$ for the symmetric GP. Using 326 training points reduces these values to $17 \times 10^{-6}\ E_h$ and $1.3 \times 10^{-6}\ E_h$ respectively. The latter RMSE is therefore less than 0.2% of the well depth, and less than 0.03% of the high-energy cutoff.

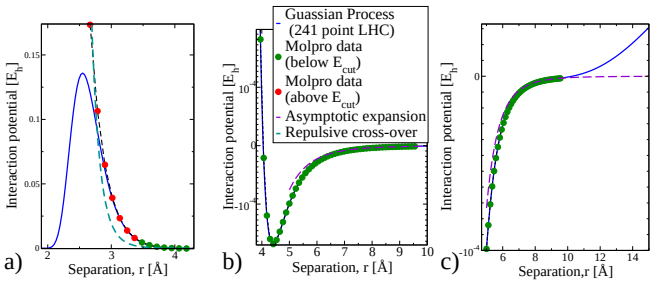

Fig. 3 CO_2-Ne Molpro calculations and the GP model, at $\cos\theta = 1$ (linear geometry), in the repulsive (a), attractive (b) and long-range (c) regions. The long-range asymptotic expansion is $E = -(0.570 + 0.182\cos^2\theta)r^{-6} - (1.704 + 7.266\cos^2\theta + 1.785\cos^4\theta)r^{-8}$.

The performance of the GP outside the training region ($E > 0.005E_h$ and $r > 8.5$Å) is shown for linear CO_2-Ne in Figure 3; results for other geometries and interactions are qualitatively similar. The extrapolation errors for points within the geometric constraint but with $E > E_{\rm cut}$ are mostly good, being a few percent or less. However, for small values of r the GP returns to its mean value of zero. This unphysical behaviour can be corrected by using a mean function with strong repulsion outside the geometric constraint. One example of the many possible choices is plotted in Figure 3(a), namely $E = E_{\rm max}\frac{1}{N}\sum_{i=1}^{N}(x_i/x_{\rm max})^{12}$, where $E_{\rm max}$ is an estimate of the typical energy at the small-r edge of the geometric constraint* and $x_{\rm max}$ is the maximum inverse distance allowed by the geometric constraint (0.67 Å^{-1} in this case). For large separations the GP tends to a small, but non-zero constant. This can be corrected for points beyond the geometric constraint, by crossing over to the long-range asymptotic expansion obtained from time-dependent perturbation theory. Figure 3(c) shows that smooth interpolation between the GP and this function will be straightforward.

4 Conclusions

The procedure described here has been used to produce intermolecular potential energy surfaces efficiently from a relatively small number of input data points. The algorithm is straightforward and easily generalised to new molecular pairs. It uses a symmetric Gaussian process, with the inverse interatomic distances as input variables. The GP is trained using energy data chosen from a Latin hypercube design, with a maximin criterion for the inverse internuclear distances.

The wide applicability and robustness of the approach has been demonstrated by testing against two systems involving CO_2, a system with a deep energy minimum (HF–HF) and a system with 48 symmetries (CH_4-N_2). In all cases the approach accurately predicts an extensive set of test data, with no *a priori* knowledge of this dataset, and gives RMSE values that are similar to, or better than, the best fits in the literature, which were generally based on thousands of training points. Furthermore, the interpolation method can be readily and directly applied to any pairwise interaction, at least for simple molecules, with no bespoke work, beyond identifying the symmetries in the system.

The approach contains three key innovations: a novel method for symmetric GP kernels; the use of inverse interatomic distances as the GP input variables; and a new strategy for positioning training data on a Latin hypercube design with a maximin criterion on the inverse intermolecular distances.

There are numerous extensions that follow from this approach. The relatively small number of training points can be used in more precise and computationally demanding potential energy calculations, then interpolated, without needing extensive test data. Application to many other chemical systems is straightforward. Furthermore, the model's accuracy against training set size could be optimised by sequentially adding training points through active learning methods[13]. This could be achieved either with or without *a priori* knowledge of the test data, depending on the nature of the potential energy data to be modelled. Another promising application is the interpolation of non-additive potentials, which are known to be difficult to fit[14]. Here the data are usually high-dimensional, vary strongly and rather unpredictably as a function of geometry, and can contain many symmetries. Finally, existing high-precision calculations could be used as training and testing data for interpolation by the algorithm. Here a sparse Gaussian process[15] could select a subset of the preexisting data on which to base computation, leading to numerically cheap, yet highly accurate, potential energy surfaces.

5 Acknowledgements

The authors are grateful to EPSRC for the award of a studentship to EU, and to the University of Nottingham for the use of the 'Minerva' high-performace computing facility.

* This can be obtained from the maximum energy, before applying the energy cut-off, over the test data (if available) or training data.

References

1 R. Chen, E. Jiao, H. Zhu and D. Xie, *Journal of Chemical Physics*, 2010, **133**, 104302.

2 H. Li, P.-N. Roy and R. J. Le Roy, *Journal of Chemical Physics*, 2010, **132**, 214309.

3 R. Hellmann, E. Bich, E. Vogel and V. Vesovic, *Journal of Chemical Physics*, 2014, **141**, 224301.

4 Y. Cui, H. Ran and D. Xie, *Journal of Chemical Physics*, 2009, **130**, 224311.

5 T.-S. Ho and H. Rabitz, *Journal of Chemical Physics*, 2000, **113**, 3960.

6 A. P. Bartok and G. Csanyi, *Int. J. Quant. Chem.*, 2015, **115**, 1051.

7 A. P. Bartok, M. C. Payne, R. Kondor and G. Csanyi, *Phys. Rev. Letts.*, 2010, **104**, 136403.

8 W. J. Szlachta, A. P. Bartok and G. Csanyi, *Phys. Rev. B*, 2014, **90**, 104108.

9 A. P. Bartok, M. J. Gillan, F. R. Manby and G. Csanyi, *Phys. Rev. B*, 2013, **88**, 054104.

10 H. J. Werner *et al.*, MOLPRO *version 2012.1: A package of ab initio programs*, http://www.molpro.net, 2012.

11 C. Rasmussen and C. K. I. William, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

12 *GPy: A Gaussian process framework in python*, http://github.com/SheffieldML/GPy, 2012–2015.

13 J. Beck and S. Guillas, *arXiv*, 2014.

14 M. T. Oakley and R. J. Wheatley, *J Chem Phys*, 2009, **130**, 034110.

15 E. Snelson and Z. Ghahramani, Advances in Neural Information Processing Systems 18, 2006, pp. 1257–1264.

Electronic supplementary information:

Interpolation of intermolecular potentials using Gaussian processes.

Elena Uteva [a], Richard S. Graham [b], Richard D. Wilkinson [c] and Richard J. Wheatley[a*].

[a] *School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.*
[b] *School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK.*
[c] *School of Mathematics and Statistics, University of Sheffield, Western Bank, Sheffield, S10 2TN UK.*

1 Latin hypercube generation

We wish to generate a dataset of model evaluations, $\{\mathbf{x}_i, f(\mathbf{x}_i)\}_{i=1}^{N}$, that can be used to train the Gaussian process, where the $\mathbf{x}_i$ represent N distinct molecular geometries. Each element of $\mathbf{x}_i$ is the inverse distance between two atoms, one from each of the molecules under consideration. The design only needs to contain points in a symmetry-distinct subspace. For example, in CO_2-Ne the O nuclei are denoted O1 and O2, and the symmetry-distinct subspace is defined such that Ne is always nearer to O1 than to O2. Space filling designs are held to be good choices for Gaussian process models, and so we will use a maxi-min criterion to evaluate candidate designs. In otherwords, we seek designs which maximise the minimum distance between any two design points. Latin hypercube (lhc) designs are used as candidate designs, as they naturally fill space to some extent, and we then choose a prefered design from a large number of candidates. We define the effective distance between points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the design to be

$$|\mathbf{x}|_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j) \tag{1}$$

and we generate a training design using the following algorithm:

- Generate a lhc in $1/r$ and rigid-body rotation angles. (For non-rigid molecules, intramolecular coordinates would also be used.)
- Convert the lhc data to atomic positions and compute all interatomic distances for pairs of atoms on separate molecules.
- Reject the geometries that don't obey the geometric constraint or lie outside the symmetry-distinct region of coordinate space.
- Reject the entire lhc if it does not contain at least the target number of geometries (usually the mean number of remaining points after the geometric constraint is applied).
- Find the minimum $|\mathbf{x}|_{ij}^2$ within the current lhc.
- Repeat for N_{it} new lhcs and return the lhc with the largest minimum $|\mathbf{x}|_{ij}^2$.

2 Symmetric covariance function

The motivating problem is modelling the H_2 - CO_2 system, which we parameterise by 6 distances:

- $r_1 = \mathrm{H}^1 \to C$
- $r_2 = \mathrm{H}^2 \to C$
- $r_3 = \mathrm{H}^1 \to O^1$
- $r_4 = \mathrm{H}^2 \to O^1$
- $r_5 = \mathrm{H}^1 \to O^2$
- $r_6 = \mathrm{H}^{\to} O^2$

The potential function f between the two molecules obeys the following symmetry relations

$$f(123456) = f(214365) = f(125634) = f(216543)$$

where $f(123456)$ denotes $f(r_1, r_2, r_3, r_4, r_5, r_6)$.

In other words, the function

$$f(x) = f(\sigma x) \forall \sigma \in K_4$$

where K_4 is the permutation group consisting of the permutations

$$\sigma_1 = (12)(34)(56), \qquad \sigma_2 = (35)(46) \qquad \sigma_3 = (12)(36)(45),$$

where we are using cyclic notation for the permutations. Note that along with the identity e, these four permutations form an abelian group that is symmetric to the Klein-4 group K_4 ($\equiv Z_2 \times Z_2$), i.e., $\sigma_i^2 = e$ and $\sigma_1\sigma_2 = \sigma_3$ etc.

2.1 A single symmetry

To start with, suppose we want to model f where f is invariant under the single permutation σ, where $\sigma^2 = e$. If we assume

$$f(x) = g(x) + g(\sigma x)$$

for some arbitrary function g, then f has the required symmetry. If we model $g(\cdot) \sim GP(0, k(\cdot,\cdot))$, then the covariance function for f is

$$k_f = \mathbb{C}\mathrm{ov}(f(x), f(x')) = k(x,x') + k(\sigma x, x') + k(x, \sigma x') + k(\sigma x, \sigma x')$$

If k is an isotropic kernel (we only actually require isotropy for each pair of vertices that swap in σ), then $k(x,x') = k(\sigma x, \sigma x')$ and $k(x, \sigma x') = k(\sigma x, x')$ as swaps only occur in pairs ($\sigma^2 = e$). So we can use

$$k_f(x,x') = k(x,x') + k(\sigma x, x')$$

saving half the computation.

2.2 Invariance under permutations in K_4

Now lets consider functions that are invariant to permutations in K_4. If we write

$$f(x) = g(x) + g(\sigma_1 x) + g(\sigma_2 x) + g(\sigma_3 x)$$

then if $g(\cdot) \sim GP(0, k(\cdot,\cdot))$

$$k_f(x,x') = k(x,x') + k(\sigma_1 x, x') + k(\sigma_2 x, x') + k(\sigma_3 x, x') + k(x, \sigma_1 x') + k(\sigma_1 x, \sigma_1 x') + \ldots k(\sigma_3 x, \sigma_3 x') \quad (2)$$

If k is isotropic, then $k(x, \sigma_i x') = k(\sigma_i^{-1} x, x')$. Thus $k(x,x') = k(\sigma_i x, \sigma_i x')$, $k(x, \sigma_i x') = k(\sigma_i x, x')$ and $k(\sigma_i x, \sigma_j x') = k(\sigma_k x, x')$ for $i \neq j \neq k$. Thus we can use

$$k_f(x,x') = k(x,x') + k(\sigma_1 x, x') + k(\sigma_2 x, x') + k(\sigma_3 x, x')$$

as a covariance function for f instead of Equation (2). This reduces the amount of compution needed to calculate the covariance functions by 75%.

Note that we don't need k to be completely isotropic for this simplification to hold, only that the covariance function is isotropic for any pair of inputs that swap in any of the permutations. So in the H_2- CO_2 system, we require the length-scales to be the same for inputs 1 and 2, and the same for inputs 3, 4, 5 and 6.