

# Can AI provide a shortcut to Slater-type orbitals?

*Elias Dalan*



UNIVERSITETET  
I OSLO

THESIS SUBMITTED FOR THE DEGREE OF  
BACHELOR OF SCIENCE IN CHEMISTRY

10 ECTS

# Contents

<b>1 Abstract</b>	<b>2</b>
<b>2 Acknowledgements</b>	<b>3</b>
<b>3 Introduction</b>	<b>4</b>
<b>4 Theory</b>	<b>5</b>
4.1 Underlying quantum chemistry . . . . .	5
4.2 Overlap and Expectation values of operators . . . . .	6
4.3 The choice of basis . . . . .	6
4.4 Gaussian process regression . . . . .	7
<b>5 Method</b>	<b>9</b>
5.1 GPR implementation . . . . .	9
5.2 Integration . . . . .	10
5.2.1 The Normalization problem . . . . .	10
5.3 Testing . . . . .	11
<b>6 Results</b>	<b>12</b>
6.1 Our cases . . . . .	12
6.2 The normalization problem . . . . .	16
6.3 Correlation between Slater-type integrals and Gaussian-type integrals . . . . .	17
6.4 Integration and regression . . . . .	19
6.4.1 Overlap integrals - 1s-like . . . . .	19
6.4.2 Kinetic and 1D-core integrals for 1s-like orbitals . . . . .	20
6.4.3 Core integrals for 2p-like orbitals . . . . .	20
6.5 Energy calculations using even tempered basis . . . . .	20
<b>7 Discussion</b>	<b>21</b>
<b>8 Conclusion and future prospects</b>	<b>23</b>
<b>9 Appendix</b>	<b>24</b>

## 1 Abstract

The aim of this project was to investigate whether or not machine learning methods could be used to predict the difference between integrals using slater-like functions and gaussian-like functions. Using scikit-learns module for Gaussian Processes (GP), regressors were trained for overlap, kinetic and potential integrals for one dimensional basis functions. To touch base with the main problem faced when doing caluclations on slater-type orbitals, the multi-center electron repulsion integrals, a three-center case consisting of two basis functions and a core-potential was also investigated. All regressors were then tested on randomized training sets which produced results in good agreement with the corresponding true values.

As a more practical implementation, the integrals predicted by the regressors were then used to solve a generalized eigenvalue problem for 1 dimensional hydrogen. The energies obtained using machine-learned integrals corresponded well with the results obtained using the true integrals.

A more personal and general goal of mine is to make computational chemistry more available to the average student. The attention span of a student is not too long, and although computations are a part of the syllabus in a lot of introductory courses to computational chemistry, students often end up spending more time setting up calculations (and waiting on them to finish!) than they do appreciating the beauty of what the computational part of chemistry has to offer. Though completely out of scope for this project, creating an environment where application and ease of use is the focus would be something I would regard as sort of an "end goal".

## **2 Acknowledgements**

Though this is not as common in bachelor theses I would like to give a thanks to some special people. Audun and Håkon, my supervisors, thank you so much for all your guidance. In a way I feel like I have learned more working with you than I have done during the previous years of my bachelor, and I thank you a lot for that. Thank you to my friends for listening to me moan about endless bugs in my code, I am truly grateful for all your support.

### 3 Introduction

Most basis sets used in modern quantum chemistry consists of basis functions constructed from Gaussian-type orbitals (GTOs). Although these functions work as a good approximation, they lack the correct behaviour at the core, and their radial decay is too high to accurately describe the electrons far away from the nucleus. Slater-type orbitals (STOs) satisfy both these conditions, however, they are difficult to work with. The difficulty with STOs arise when doing three- and four-center integrals in higher dimensions, so if a machine learning model could aid in solving two- and three-center cases in 1D there is a probability that the model can be translated to higher dimensions and more complex systems, which would greatly reduce the computational time when using basis sets consisting of STOs [4]. Machine learning in general has in recent years become an intrinsic part of modern science and is continuing to develop as a tool in the scientists toolbox. Amongst the many various approaches in quantum chemistry, the so called  $\Delta$ -ML method [2] has been used extensively to correct energy and property calculations of conventional low-level methods with only minor modifications of existing codes. Within this context we find an interesting question; can ML-methods correct the Gaussian-based integrals into their Slater-type counterparts? This question will likely have many answers, and this text documents my attempt at providing answers and more information on this topic.

The thesis is composed as follows: The reader is in the next part introduced to the theory of this project. Necessary topics regarding quantum chemistry are presented first, followed by an introduction to the mathematical expressions we have worked with in this thesis. A description of the two main stars, GTOs and STOs are provided, followed by an explanation of the machine learning theory one needs to be able to follow this project. The theory is described at a level which is meant to be suited for a third-year bachelor student.

The method section describes the practical part of this project, and outlines the scope and limitations of our ML-model, in addition to an explanation of how our regressors were tested, and how we model their accuracy.

The results are then presented, followed by a discussion and a brief conclusion.

## 4 Theory

### 4.1 Underlying quantum chemistry

In quantum chemistry, observable quantities like position, momentum and energy are represented by Hermitian operators. We will only be concerned with one operator (or three, depending on how you look at it), namely the Hamiltonian. The Hamiltonian  $\hat{H}$  is defined as the sum of the kinetic energy operator  $\hat{T}$  and the potential energy operator  $\hat{V}$  and its corresponding observable is energy. The energy of a system governs the properties the system might have, such as reactivity and stability. The Hamiltonian for a particle in a potential  $V$  is given by:

$$\hat{H} = \hat{T} + \hat{V} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}, \mathbf{t}), \quad (1)$$

where  $\hat{T}$  is the kinetic operator and  $\hat{V}$  is the potential operator. Quantum chemistry is to a large extent made possible by assuming fixed nucleis, the so called Born-Oppenheimer approximation, and typically no external field. Under these conditions the wavefunction is separable in time and space, and a great deal of chemical knowledge can be obtained by only solving the time-independent Schrödinger equation (TISE). The time-independent Hamiltonian and the energy of a system are related through the TISE,

$$\hat{H}|\psi_k\rangle = E_k|\psi_k\rangle. \quad (2)$$

The time-independent Schrödinger equation is an eigenvalue problem, which for a given set of  $\{\psi_k\}$  is solved to obtain the corresponding eigenstates  $\psi_k$  and eigenvalues  $E_k$ . In general the eigenvalues of a Hermitian operator is interpreted as the possible outcomes a measurement of the corresponding observable can yield. The eigenstates are interpreted as the state which the system collapses to after a measurement of it corresponding eigenvalue. For the Hamiltonian the measurement yields the energy of the state, and the system then collapses to the corresponding energy-eigenstate.

An important theorem to consider is the variational principle. For a given  $\hat{H}$  there exists some exact solutions to the Schrödinger equation, shown as  $\psi_k$  in Equation 2. Given some trial function  $|\psi_{trial}\rangle$  it can be shown that

$$\frac{\langle\psi_{trial}|\hat{H}|\psi_{trial}\rangle}{\langle\psi_{trial}|\psi_{trial}\rangle} \geq E_0, \quad (3)$$

always holds, as long as the function satisfies the boundary conditions of the system we want it to describe [10]. The equality holds only if  $|\psi_{trial}\rangle = |\psi_0\rangle$ . Quantum chemistry would be impossible without approximations, and the variational principle allows us to come up with good approximations of the ground state. If our wavefunctions depend on some parameters  $\mathbf{c}$  we can vary these parameters to minimize the energy and measure the quality of our approximation based on how close the energy is to the real ground state! To be able to evaluate expression 3 we have to evaluate the following expressions:

$$\langle\psi|\hat{H}|\psi\rangle, \langle\psi|\psi\rangle \quad (4)$$

These expressions are defined in section 4.2

## 4.2 Overlap and Expectation values of operators

The two main mathematical expressions we will be concerned about are overlap integrals and the expectation values of operators (though overlap integrals can be viewed as the expectation value of the identity operator  $\mathbb{1}$ ). The overlap between two functions,  $S_{mn}$ , is given by:

$$S_{mn} = \langle \psi_m | \psi_n \rangle = \int \psi_m^* \psi_n \, d\mathbf{r} \quad (5)$$

The asterisk on  $\psi_m$  denotes the function's complex conjugate. The expectation values of some operator  $\hat{\Omega}$  is given by:

$$\Omega_{mn} = \langle \psi_m | \hat{\Omega} | \psi_n \rangle = \int \psi_m^* \hat{\Omega} \psi_n \, d\mathbf{r} \quad (6)$$

Based on your choice of  $\psi$  these expressions can be hard to evaluate analytically (or even impossible), so one often has to resort to numerical methods, especially when working in higher dimensions. Numerical methods are more computationally expensive, and take more time, so a balance between precision and time management has to be found. This leads us to the choice of basis, discussed in the next section.

## 4.3 The choice of basis

One of the most important choices one has to make when doing some sort of computational chemistry is the choice of basis. One could, in theory, construct a complete basis set by including an infinite amount of functions, but this is obviously not practically possible. Our choice of basis is made with the underlying physics in mind: What properties does our system have? Which properties are essential, and what can be approximated to decrease computational complexity? Most modern quantum chemistry code use a basis of GTOs,

$$\psi_{ijk}^{GTO} = N_{GTO} x^i y^j z^k e^{-\alpha r^2}. \quad (7)$$

According to Kato's cusp condition [6], an electronic wave function is supposed to have a cusp at the center, to mimic the electrons behaviour near the core. GTOs lack this property, as they are smooth and continuous at the centre. A more preferable choice is Slater-type orbitals (STOs)[3][8],

$$\psi_{ijk}^{STO} = N_{STO} |\mathbf{r}|^{l-(i+j+k)} x^i y^j z^k e^{-\alpha |\mathbf{r}|}, \quad (8)$$

since they satisfy this condition.  $N_{GTO}$  and  $N_{STO}$  are the normalization constants for the GTO and STO wavefunctions,  $i, j$  and  $k$  are quantum numbers governed by the total angular momentum  $l$ , where  $l = i + j + k$ .  $\alpha$  controls the width of the functions (large alpha gives a tight function, while a smaller alpha gives a more diffuse function).

In addition to being able to describe the cusp at the core, STOs also decay slower in terms of  $\mathbf{r}$  and therefore allow better representation of electrons further from the nucleus [4]. Assuming that we can treat all atoms as perturbed versions as Hydrogen-like atoms and additionally that any molecule can be regarded as a perturbed version of collections of these Hydrogen-like atoms, STOs immediately come to mind as a suitable basis. If STOs are so great, why are

GTOs so prevalent in quantum chemistry? The answer lies in computational efficiency. Integrals with Gaussian-type orbitals are much easier to evaluate than their Slater counterparts, and by approximating an STO using multiple GTOs, like with the STO-nG basis sets, you can make up for loss in accuracy [4].

#### 4.4 Gaussian process regression

A major problem when doing regression on some dataset where the corresponding output does not necessarily have a closed form (ex. a quadratic function) is choosing the appropriate approximation. When fitting some data to an non ambiguous function there is an underlying restriction, also called an inductive bias, imposed by the model function. GPR works around this problem by letting the data speak more for itself. A thorough presentation of the underlying theory is out of scope of this project, but a short summary will be provided. A full description can be found in [7]. The principle assumption of Gaussian processes (GPs) is that our data is distributed as a multivariate Gaussian distribution. This means that GPR does not return a single value as an output, but rather a probability distribution. Before we go further we will have to define two expressions, the kernel and the covariance.

The first is the kernel function, often denoted as  $k(x_i, x_j)$ . The kernel describes the covariance between two inputs, and should be chosen to best fit the data you are trying to predict. The optimal choice would obviously be to use a kernel specifically designed for the system you are trying to model, but creating such a kernel is time consuming, and might not even be worth the time.

We define the covariance-matrix  $K$  as the matrix-representation of the kernel given some input data  $\mathbf{v} = (v_1, v_2, v_3, \dots, v_m)$  and  $\mathbf{w} = (w_1, w_2, w_3, \dots, w_n)$  (These can also be matrices, but vectors were chosen for ease of visualization), in other words:

$$K(\mathbf{v}, \mathbf{w}) = \begin{bmatrix} k(v_1, w_1) & k(v_1, w_2) & k(v_1, w_3) & \dots & k(v_1, w_n) \\ k(v_2, w_1) & k(v_2, w_2) & k(v_2, w_3) & \dots & k(v_2, w_n) \\ k(v_3, w_1) & k(v_3, w_2) & k(v_3, w_3) & \dots & k(v_3, w_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k(v_m, w_1) & k(v_m, w_2) & k(v_m, w_3) & \dots & k(v_m, w_n) \end{bmatrix} \quad (9)$$

Through the principle assumption of GPs one can arrive at an expression for the best prediction  $\mathbf{y}_*$  to some input labels  $X_*$ , given some training labels  $X$  and the corresponding output features  $\mathbf{y}$ , and is given by

$$\mathbf{y}_* = K(X, X_*)K(X, X)^{-1}\mathbf{y}. \quad (10)$$

A major strength of this type of modeling is that we can obtain an expression for the uncertainty of our prediction, measured by its variance,

$$var(\mathbf{y}_*) = K(X_*, X_*) - K(X, X_*)K(X, X)^{-1}K(X, X_*)^T. \quad (11)$$

An example of a GPR used to fit  $f(x) = e^{-(x-1)^2} + e^{-(x+1)^2}$  is shown in Figure 1.

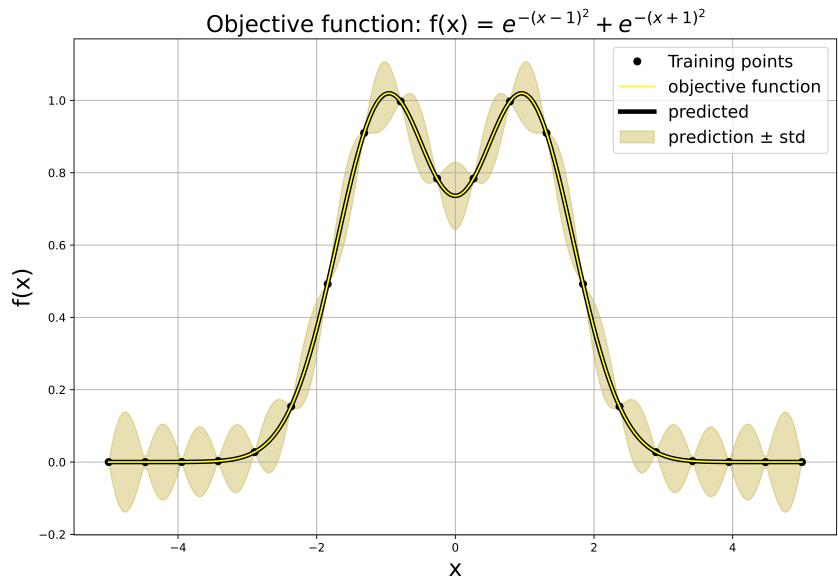


Figure 1: GPR fit to a uniform grid of 20 sets of  $(x, f(x))$  for  $x \in [-5, 5]$ .

## 5 Method

All models were implemented using atomic units. That is, setting  $m_e = e = \hbar = 4\pi\varepsilon_0 = 1$ , where  $m_e$  is the electron mass,  $e$  is the elementary charge,  $\hbar$  is the reduced planck constant and  $4\pi\varepsilon_0$  is the Couloumb constant.

GPRs were trained to predict the difference between some type of integral using Slater-like integrand  $I(\chi_{STO})$  and the corresponding Gaussian-type integrand  $I(\chi_{GTO})$ ,

$$\Delta I = \int I(\chi_{STO}) dx - \int I(\chi_{GTO}) dx. \quad (12)$$

### 5.1 GPR implementation

GPR was implemented in Python using Scikit-Learns module for Gaussian processing [7]. As a warm up for this project I implemented a simple GPR module, which can be found here [1]. The kernel chosen was the radial basis function (RBF)

$$RBF(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)}{2l^2}\right), \quad (13)$$

where  $d(x_i, x_j)$  is the Euclidean distance between  $x_i$  and  $x_j$  and  $l$  is called the length-scale parameter. The parameter  $l$  can be adjusted so that the covariance-matrix better represents our data. There was not done much optimization for the length-scale parameter, as this is very time consuming using exhaustive search, but a few tests were done and the first length scale parameter that let the fitting process converge was used. All regressors were trained on all possible combinations of our parameters. In other words: The regressors trained on two parameters were trained on a uniform quadratic grid, while the regressors trained on three were trained on a uniform cubic grid. The training boundaries for the different regressors are shown in Table 1.

1s overlap integrals:		
Parameter	Minimum value value	Maximum value
$\alpha$	0.1	5
$\beta$	0.1	5
$x_0$	0	5
1s kinetic integrals:		
Parameter	Minimum value	Maximum value
$\alpha$	0.1	5
$\beta$	0.1	5
1s core integrals:		
Parameter	Minimum value	Maximum value
$\alpha$	0.1	5
$\beta$	0.1	5
2p core integrals		
Parameter	Minimum value	Maximum value
$x_a$	-0.1	-5
$x_b$	0.1	5
$x_c$	-2	2

Table 1: Parameter boundaries for the different regressors.

## 5.2 Integration

Analytical expressions were obtained for most of the integrals, which are shown in the Appendix. All analytical expressions were obtained using Wolfram Alpha [5]. Wolfram Alpha did not give an analytical expression for different center overlap integrals, and numerical integration using Scipy quadrature-function was used [11]. The integrals for the 2p system was implemented using the Braketlab environment [9]. Integrals are done numerically using this package, and 1-dimensional integrals are evaluated using scipys quadrature-module. An error-estimate of less than  $10^{-9}$  was obtained for all integrals using numerical methods, which was found to be satisfactory.

### 5.2.1 The Normalization problem

If one is to interpret a wavefunction  $\Psi$  as a probability amplitude, and therefore the square magnitude as a probability density, one has to tackle the task of normalizing it. That is, finding a normalization constant  $N$  such that  $\langle \Psi' | \Psi' \rangle = 1$ , where  $\Psi' = N\Psi$ . In order to ensure that the full wavefunction is normalized, it is convenient to use normalized basis functions. The simplest (1S) Gaussian-like and Slater-like functions are

$$\psi_{GTO} = e^{-\alpha x^2} \text{ and } \psi_{STO} = e^{-\alpha|x|}, \quad (14)$$

we calculate the overlap integrals:

$$\langle \psi'_{GTO} | \psi'_{GTO} \rangle = N^2 \int_{\mathbb{R}} e^{-\alpha x^2} e^{-\alpha x^2} dx \quad (15)$$

$$= N^2 \sqrt{\frac{\pi}{2\alpha}} \quad (16)$$

And for the STO-like functions.

$$\langle \psi'_{STO} | \psi'_{STO} \rangle = N^2 \int_{\mathbb{R}} e^{-\alpha|x|} e^{-\alpha|x|} dx \quad (17)$$

$$= \frac{N^2}{\alpha} \quad (18)$$

The normalization constant is given by,

$$N = \sqrt{\frac{1}{\langle \Psi' | \Psi' \rangle}}. \quad (19)$$

By inserting the overlap integrals into Equation 20 we obtain the two normalization constants:

$$N_{GTO} = \sqrt[4]{\frac{2\alpha}{\pi}}, \quad N_{STO} = \sqrt{\alpha} \quad (20)$$

An interesting point to consider is whether or not our network should be trained on normalized integrals. Though this might not seem important, introducing normalization constants will change the labels that our network is trying to predict. If the case is that normalization introduces some regularity in our data, then the regressor will have an easier time predicting the labels. On the other hand it might make the mathematical model we are trying to predict more complex, which will result in poorer  $\Delta I$  predictions. Another point to consider is whether or not the inputs should be scaled. This will not be done, but is discussed in **Conclusion and future prospects**.

### 5.3 Testing

The GPRs were all tested on 10000 sets of their respected parameters that were randomly generated (inside of the training domain of course). The regressors predictions were compared to the true outputs, and the average error,

$$\bar{E} = \frac{1}{N} \sum_{i=1}^N |y_i - y_{*i}|, \quad (21)$$

the maximum error,

$$\max E = \max |y - y_*|, \quad (22)$$

and the  $R^2$  norm,

$$R^2 = 1 - \frac{\sum_i (y_i - y_{*i})^2}{\sum_i (y_i - \bar{y})^2}, \quad (23)$$

were used to determine the quality of the regressors.

## 6 Results

### 6.1 Our cases

Three different one-dimensional systems were investigated:

- Overlap between 1s-like orbitals, where one is centered at zero and the other is centered at  $x_0$ . They both had varying exponents  $\alpha$  and  $\beta$ .
- Kinetic and potential integrals between two 1s-like orbitals using the 1D-Hamiltonian, with  $V(x) = -\frac{1}{\sqrt{x^2+1}}$ . Both orbitals were centered at zero, with varying exponents  $\alpha$  and  $\beta$ .
- Three-center integrals between two 2p-like orbitals and a core-potential. The two basisfunctions and the potential was free to move in between specified intervals.

The different integrals that were calculated are shown in Equations 24, 25, 26 and 27. They are shown as their unnormalized Slater-type variants.

$$\langle \psi_\alpha | \psi_{\beta x_0} \rangle = \int_{-\infty}^{\infty} e^{-\alpha|x|} e^{-\beta|x-x_0|} dx \quad (24)$$

$$\langle \psi_\alpha | \hat{T} | \psi_\beta \rangle = \int_{-\infty}^{\infty} -e^{-\alpha|x|} \frac{1}{2} \frac{d^2}{dx^2} e^{-\beta|x|} dx \quad (25)$$

$$\langle \psi_\alpha | \hat{V} | \psi_\beta \rangle = \int_{-\infty}^{\infty} -e^{-\alpha|x|} \frac{1}{\sqrt{x^2+1}} e^{-\beta|x|} dx \quad (26)$$

$$\langle \psi_{x_a} | \hat{V}_{x_c} | \psi_{x_b} \rangle = \int_{-\infty}^{\infty} -(x - x_a) e^{-|x-x_a|} \frac{1}{\sqrt{(x - x_c)^2 + 1}} (x - x_b) e^{-|x-x_b|} dx \quad (27)$$

Analytical expressions can be obtained for some of these expressions, both for the Slater-variants and for the Gaussian variants. These are listed in the Appendix. Figures 2, 3 and 4 show the three different cases that were investigated.

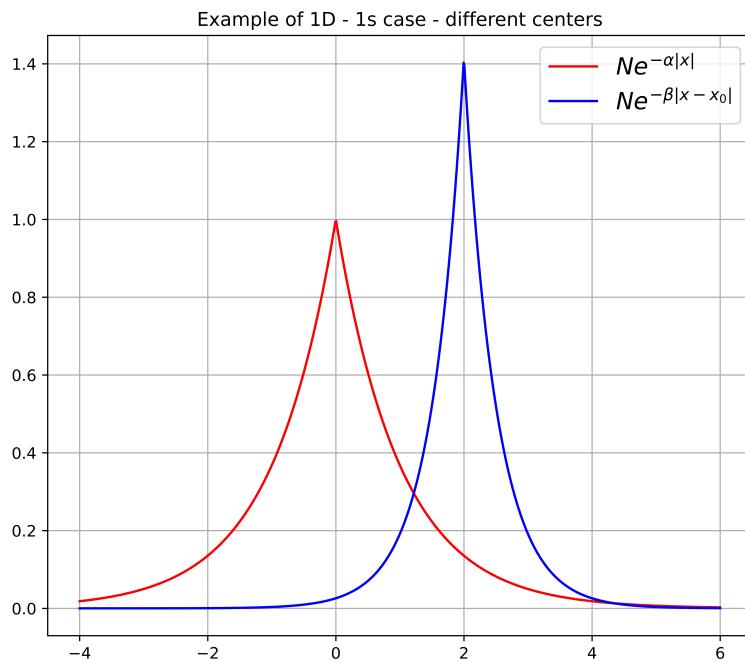


Figure 2: Example of the 1-dimensional 1s case. Here  $\alpha = 1, \beta = 2, x_0 = 2$ .

The first calculations done was overlap between a system of two 1s-like hydrogen orbitals. The expression for the overlap is shown in Equation 24 and an example of this system is shown in Figure 2.

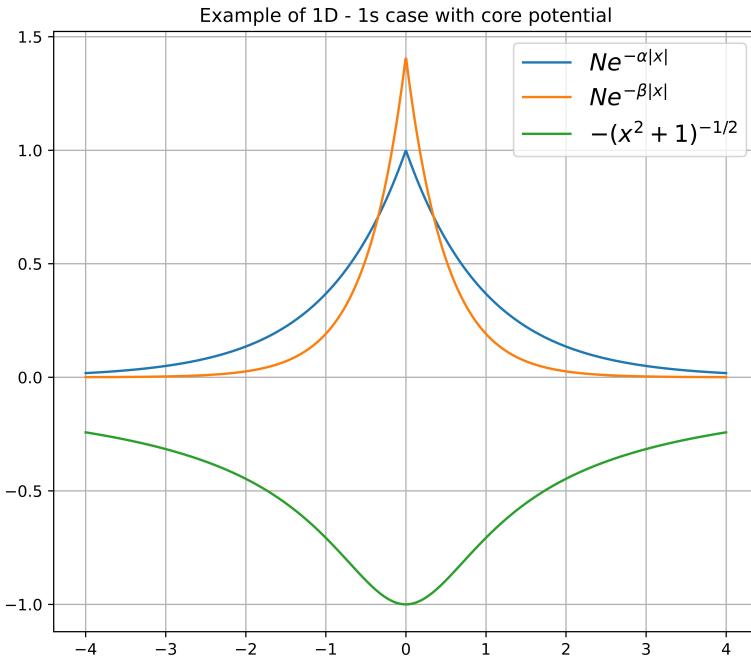


Figure 3: Example of the same-center 1s system with a Hydrogen-like core potential.  $\alpha = 1, \beta = 2$ .

Second system investigated was a same-center system of two 1s-like hydrogen orbitals, in addition to a Hydrogen core-potential. On this system we calculated the kinetic and potential integrals, shown in Equations 25 and 26. An example of this system is shown in 3.

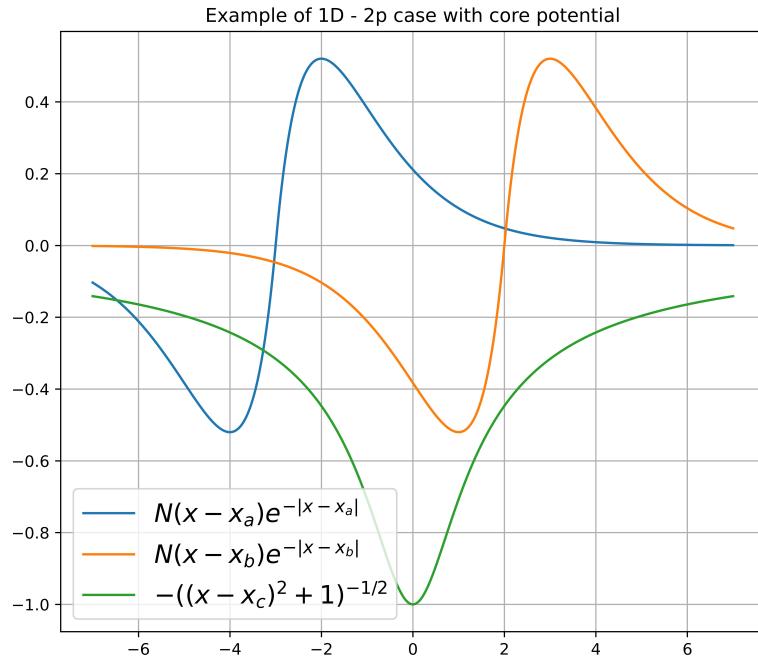


Figure 4: Example of a three-center 1D system of two p-like orbitals and a Hydrogen-like core potential.  $x_a = -3, x_b = 2, x_c = 0$ .

The last system investigated was a three-center system consisting of two p-like orbitals and a hydrogen-like core potential. An example is shown in Figure 4. The expression for the core-integral is shown in Equation 27.

## 6.2 The normalization problem

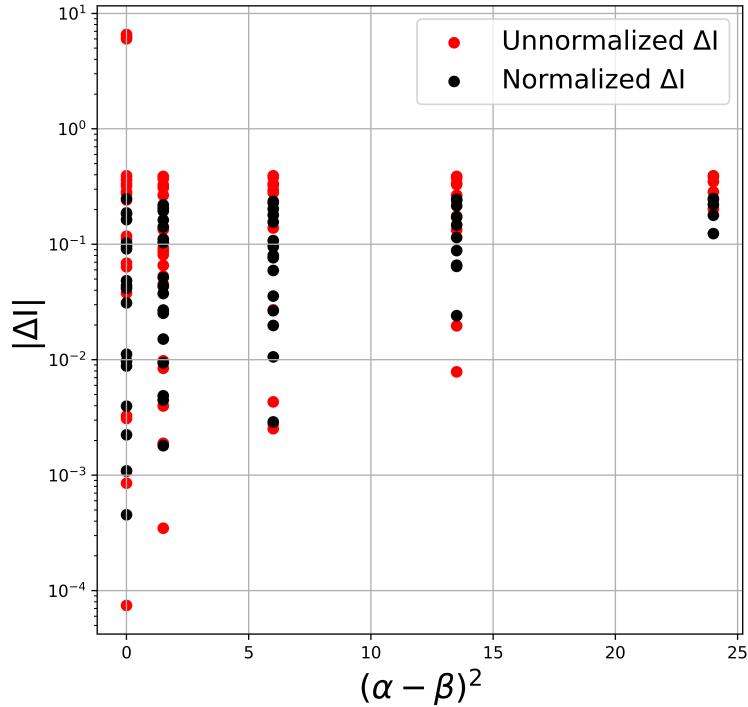


Figure 5: Comparison of normalized vs unnormalized difference between overlap integrals  $\langle \psi_\alpha | \psi_{\beta x_0} \rangle$ .

Figure 5 shows  $\Delta I = \langle STO_\alpha | STO_{\beta x_0} \rangle - \langle GTO_\alpha | GTO_{\beta x_0} \rangle$ . The unnormalized  $\Delta I$ s generally span a larger range, especially for lower differences between  $\alpha$  and  $\beta$ .

Regressor	$N_{\text{parameter}}$	Unnormalized	Normalized
Overlap - 1s	10	0.954	0.990
Kinetic - 1s	20	0.999997	0.999999
Core - 1s	20	0.99992	0.99999
Core - 2p	10	0.971	0.989

Table 2:  $R^2$  score for different regressors

The  $R^2$  scores clearly shown an increase for normalized integrals.

### 6.3 Correlation between Slater-type integrals and Gaussian-type integrals

The integrals used in the correlation plots were set up using the same grids used in section 5.4.

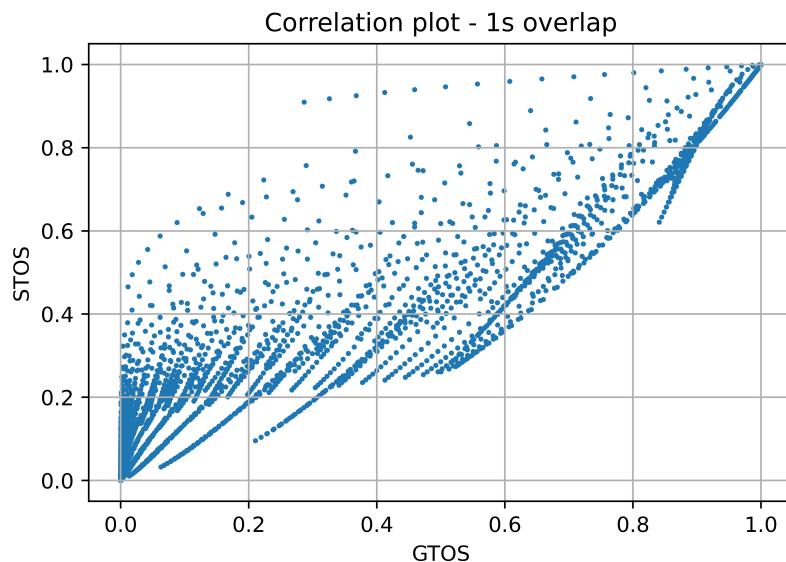


Figure 6: Correlation plot of  $\langle STO_{\alpha} | STO_{\beta x_0} \rangle$  and  $\langle GTO_{\alpha} | GTO_{\beta x_0} \rangle$ .

Figure 6 is pretty chaotic, but some correlation can be seen, as groups of points come together to form lines.

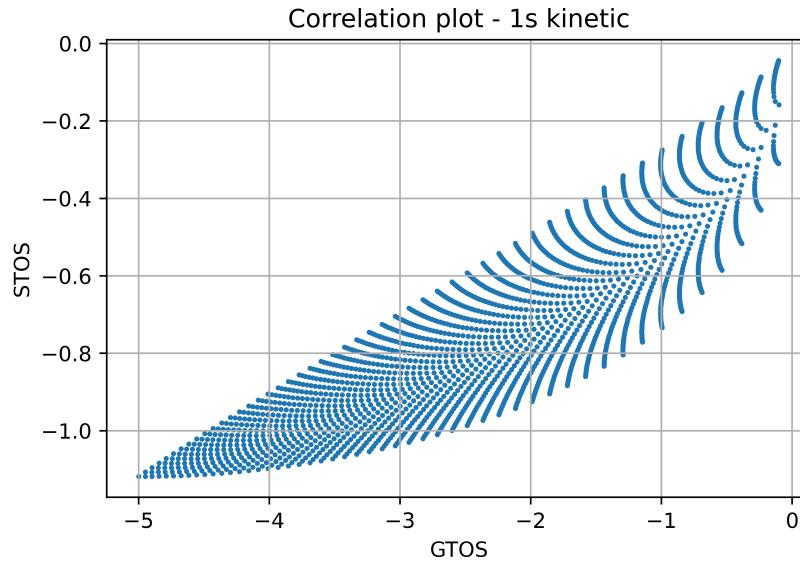


Figure 7: Correlation plot of  $\langle STO_\alpha | \hat{T} | STO_\beta \rangle$  and  $\langle GTO_\alpha | \hat{T} | GTO_\beta \rangle$ .

Figure 7 shows clear relation between the Slater-like and Gaussian-like integrals. They form something that almost looks like a leaf, possibly the Barnsley fern.

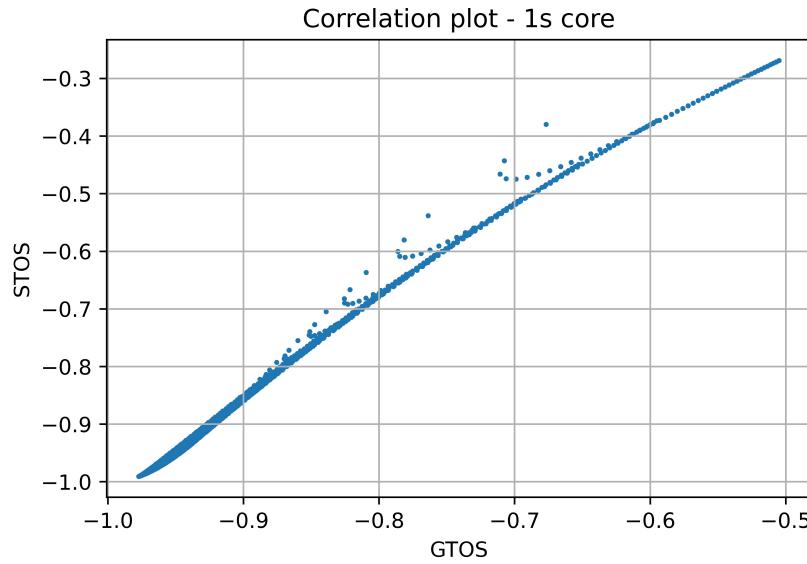


Figure 8: Correlation plot of  $\langle STO_\alpha | \hat{V} | STO_\beta \rangle$  and  $\langle GTO_\alpha | \hat{V} | GTO_\beta \rangle$ .

The correlation plots of the 1s core integrals seem to have an almost linear relationship, with the exception of some outliers.

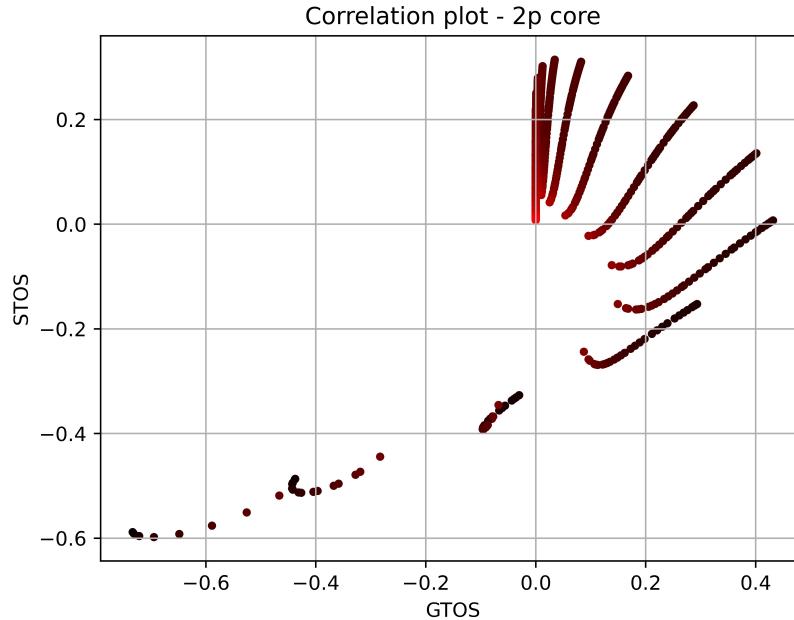


Figure 9: Correlation plot of  $\langle STO_{x_a} | \hat{V}_{x_c} | STO_{x_b} \rangle$  and  $\langle GTO_{x_a} | \hat{V}_{x_c} | GTO_{x_b} \rangle$ . The color gradient is proportional to  $|x_a - x_c| + |x_b - x_c|$ , fading from red to black as the value increases.

Though the correlation plot is a bit all over the place, there is a clear relationship between the two integrals, looking almost like branches of a tree. Each branch corresponds to a certain value of  $|x_a - x_b|$ .

## 6.4 Integration and regression

### 6.4.1 Overlap integrals - 1s-like

$N_{\alpha,\beta,x_0}$	$N_{grid}$	Integration time (s)	GPR setup time (s)	Max error	Mean error
10	1000	3.20	2.20	0.182	0.0103
15	3375	11.4	82.5	0.110	0.00302
20	8000	27.6	756	0.0665	0.000947

Table 3: Results for overlap integrals

#### 6.4.2 Kinetic and 1D-core integrals for 1s-like orbitals

$N_{\alpha,\beta}$	$N_{grid}$	Integration time (ms)	GPR setup time (s)	Max error	Mean error
10	100	0.199	0.00433	0.0301	0.00292
25	625	0.399	0.613	0.00364	$3.81 \cdot 10^{-5}$
50	2500	1.99	13.5	0.000301	$1.63 \cdot 10^{-6}$
80	6400	4.39	289	0.000107	$1.16 \cdot 10^{-6}$

Table 4: Results for Kinetic integrals

$N_{\alpha,\beta}$	$N_{grid}$	Integration time (ms)	GPR setup time (s)	Max error	Mean error
10	100	0.110	0.0478	0.0183	0.000532
25	625	4.39	0.822	0.00128	$5.28 \cdot 10^{-5}$
50	2500	27.6	35.7	0.000228	$7.41 \cdot 10^{-6}$
80	6400	67.9	184	$9.32 \cdot 10^{-5}$	$3.22 \cdot 10^{-6}$

Table 5: Results for 1D core integrals

#### 6.4.3 Core integrals for 2p-like orbitals

$N_{A,B,x_0}$	$N_{grid}$	Integration time (s)	GPR setup time (s)	Max error	Mean error
10	1000	62.6	4.98	0.0665	0.000233
15	3375	207	104	0.00201	$4.37 \cdot 10^{-6}$
20	8000	491	596	$4.61 \cdot 10^{-5}$	$1.77 \cdot 10^{-6}$

Table 6: Results for Core integrals 2p

### 6.5 Energy calculations using even tempered basis

The time-independent Schrödinger equation for 1D Hydrogen was solved, using the potential

$$V(x) = -\frac{1}{\sqrt{x^2 + 1}}, \quad (28)$$

using an even tempered basis [4] consisting of 4 simple STOs with exponents  $\zeta = \{\alpha, \alpha\beta, \alpha\beta^2, \alpha\beta^3\}$ ,  $\alpha = 1, \beta = 1.2$ . The results for both normal integration methods and using machine learned integrals are shown in table 7, in addition to the relative error,  $E_{rel} = \frac{|E_i - E_{*i}|}{|E_i|}$ .

State	Normal integrals	ML integrals	relative error
$E_0$	-0.66976181	-0.66970605	$8.32 \cdot 10^{-5}$
$E_1$	-0.07072133	-0.08134634	0.150
$E_2$	1.29311439	1.08375565	0.162
$E_3$	21.8475332	18.12355601	0.170

Table 7: Results for energy calculations.

The relative error for the ground state is of the same order as the maximum error for integration reported in Tables 3, 4, 5 and 6. The relative error of the excited states are of higher magnitude than for the ground state.

## 7 Discussion

We settled upon three cases to investigate, as described in Figures 2, 3 and 4. By limiting ourselves to 1 dimensional cases, most same-centered 1S integrals can pretty easily be obtained as provided in the Appendix. We still want to maintain generality, however, as analytical evaluation of three-center integrals are unobtainable even in one dimension, and thus connect directly to the main challenge of using an STO-basis: multicenter electron repulsion integrals. If our procedure works for one dimensional three-center nuclear repulsion integrals there is a strong indication that it would also work in the three dimensional cases.

For the three cases described, it was easy to investigate the correlation between the GTOs and STOs as a preliminary test. There is no point in trying to build a machine learning model if the correlation of GTOs and STOs in our cases are completely random. The regularity shown in Figure 6, 7, 8 and 9 shows clear and unique relationships between the integrals for all cases, and is worth further investigation.

In order to investigate normalization, regressors were trained on both normalized and unnormalized integrals for all cases. As shown in Table 2 and Figure 5 normalization clearly increases the precision of our regressors. In addition to this normalization seems to have a much greater impact on overlap integrals, while the effect is much smaller on kinetic and core integrals. This is suspected to be because of two reasons:

1.  $\langle \psi_a | \psi_b \rangle = 1$  for any and all  $\psi_a = \psi_b$ .
2.  $\langle \psi_a | \hat{\Omega} | \psi_a \rangle \neq \langle \psi_b | \hat{\Omega} | \psi_b \rangle$ .

1) is always true for normalized integrals, which means that a regularity is introduced in our model, which in turn makes it easier to learn for a regressor. This regularization is only true for  $\hat{\Omega} = \mathbf{1}$ , which means that normalization cannot necessarily be expected to improve a GPR for any given operator. This gives some great insight into how the learnability of  $\Delta I$  can be impacted by transformation of the functional form of our data. This flexibility shows that, in general, both normalization and an optimal transformation of the exponent between the GTOs and STOs has to be done to arrive at a minimization of the error in  $\Delta I$ .

From the correlation plots we can conclude that there exists some correlation between the Slater- and Gaussian-type integrals, which is essential if we want to create some kind of machine-learning model. A question that is natural to ask is whether or not the RBF kernel is the right kernel to depict this correlation through covariance. The principle assumption of the RBF kernel is that  $d(x_i, x_j)$  governs the covariance, but this is not necessarily the case, especially for integrals between varying exponents. The degree of correlation should be different between neighbouring points on the short scale and the large scale, but this is not the case when using the standard RBF kernel.

All systems show continuous decrease in error for an increasing set of grid points, which is to be expected. The overlap integrals have the highest error of the cases investigated, but this is not surprising with the correlation plot in Figure 6 in mind. The correlation plots for both the Kinetic case and the Hydrogen core-potential case show clear signs of an underlying correlation, while

the overlap case is much more complex. The three-center plot can be deceiving, as there is not as much continuity as there is in the other plots, but as shown by the color gradient there is a clear relation between similar systems. The other cases hover around a max error of  $10^{-4}$  for the biggest grids which is quite a ways off from what would be preferable for quantum computations. Considering that no optimization of the regressor was performed, these results are rather encouraging. The biggest challenge our regressors face are computational time during integration and setup, especially when increasing the grid size. As shown in Equation 10 we need to compute the inverse of the covariance matrix, and inverting a  $n \times n$  matrix generally scales like  $O(n^3)$ , so increasing the grid size greatly increases time spent on both integration and regressor setup.

The error of the three-center case is roughly of the same magnitude as the more trivial cases. Although the dependence on higher dimensionality is yet to be determined, this indicates that the procedure can work on non-trivial multi-center cases.

For the energy-calculation the ground state error is in good agreement with the integration errors, which is promising in regards to any sort of practical implementation of this model.

## 8 Conclusion and future prospects

The implementations gave accurate results, which can be systematically improved by increasing the resolution of the training grid. The results of energy-calculations gave good results for the ground state, which means that this model is well suited for optimization using the variational principle. This is not a definite proof that the model will be able to properly extend to higher dimensions while maintaining good precision, but it is a good indicator.

The results for three-center integrals where  $\Delta I = \Delta I(x_a, x_b, x_c)$  gives hope for application of this model doing three- and four-center electron repulsion integrals in higher dimensions, which is where an application could really be a breakthrough. The major problem in higher dimensions arise from computational complexity, so the extensions will have to maintain the same level of accuracy to make up for the integration and setup time.

The GPR models developed here could easily be integrated into existing quantum chemistry code, as a simple layer between the integrator and the iterative solvers, e. g. in Hartree-Fock or DFT solvers. If the error estimated of our regressor are lower than the system we are integrating it into, the error estimates can be used as error bounds for the final results obtained.

The next step would be to extend the model to higher dimensions, and to investigate how prepossessing, scaling and transformations the inputs can affect the GPRs. It is of utmost interest to improve the kernel function, to more accurately describe covariance between exponential functions. If GPRs are shown to not be a suitable model for more complex systems the regressor could easily be substituted for something more fit.

## 9 Appendix

All analytical expressions were obtained using Wolfram Alpha [5].

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} e^{-\beta x^2} dx = \sqrt{\frac{\pi}{\alpha + \beta}} \quad (29)$$

$$\int_{-\infty}^{\infty} e^{-\alpha|x|} e^{-\beta|x|} dx = \frac{2}{\alpha + \beta} \quad (30)$$

$$\int_{-\infty}^{\infty} e^{-\alpha x^2} e^{-\beta(x-x_0)^2} dx = e^{-\frac{\alpha\beta x_0^2}{\alpha+\beta}} \sqrt{\frac{\pi}{a+b}} \quad (31)$$

$$-\frac{1}{2} \int_{-\infty}^{\infty} e^{-\alpha x^2} \frac{d^2}{dx^2} e^{-\beta x^2} dx = \alpha\beta \sqrt{\frac{\pi}{(\alpha + \beta)^3}} \quad (32)$$

$$-\frac{1}{2} \int_{-\infty}^{\infty} e^{-\alpha|x|} \frac{d^2}{dx^2} e^{-\beta|x|} dx = \frac{(\alpha\beta)^2}{\alpha + \beta} \quad (33)$$

$$-\int_{-\infty}^{\infty} e^{-\alpha x^2} \frac{1}{\sqrt{x^2 + 1}} e^{-\beta x^2} dx = -e^{\frac{\alpha+\beta}{2}} K_0\left(\frac{\alpha + \beta}{2}\right) \quad (34)$$

Where  $K_0$  is the modified Bessel function of the second kind of order 0.

$$-\int_{-\infty}^{\infty} e^{-\alpha|x|} \frac{1}{\sqrt{x^2 + 1}} e^{-\beta|x|} dx = -\pi(H_0(\alpha + \beta) - Y_0(\alpha + \beta)) \quad (35)$$

Where  $H_0$  is the struve function of order 0 and  $Y_0$  is the Bessel function of second kind of order 0.

## References

- [1] Elias Dalan. Gaussian process regression tools.
- [2] Pavlo O. Dral. Quantum chemistry in the age of machine learning. *The Journal of Physical Chemistry Letters*, 11(6):2336–2347, 2020. PMID: 32125858.
- [3] J. Fernández Rico, R. López, A. Aguado, I. Ema, and G. Ramírez. Reference program for molecular calculations with slater-type orbitals. *Journal of Computational Chemistry*, 19(11):1284–1293, 1998.
- [4] Jensen Frank. *Introduction to Computational Chemistry*. Wiley, 2007.
- [5] <https://www.wolframalpha.com/>. Wolfram research.
- [6] Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2):151–177, 1957.
- [7] Eric Schulz, Maarten Speekenbrink, and Andreas Krause. A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85:1–16, 2018.

- [8] David Sherill. Basis sets part 1, 2021.
- [9] Audun Skau Hansen. Braketlab.
- [10] A. Szabo and N.S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Books on Chemistry. Dover Publications, 1996.
- [11] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.