arXiv:1306.1812v1 [physics.chem-ph] 7 Jun 2013

Orbital-free Bond Breaking via Machine Learning

John C. Snyder,[1] Matthias Rupp,[2] Katja Hansen,[3] Leo Blooston,[4] Klaus-Robert Müller,[5,6] and Kieron Burke[1]

[1] *Departments of Chemistry and of Physics, University of California, Irvine, CA 92697, USA*
[2] *Institute of Pharmaceutical Sciences, ETH Zurich, 8093 Zürich, Switzerland*
[3] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany*
[4] *Department of Chemistry, University of California, Irvine, CA 92697, USA*
[5] *Machine Learning Group, Technical University of Berlin, 10587 Berlin, Germany*
[6] *Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea*

(Dated: June 10, 2013)

Machine learning is used to approximate the kinetic energy of one dimensional diatomics as a functional of the electron density. The functional can accurately dissociate a diatomic, and can be systematically improved with training. Highly accurate self-consistent densities and molecular forces are found, indicating the possibility for ab-initio molecular dynamics simulations.

Kohn-Sham density functional theory (KS-DFT) [1, 2] is a widely used electronic structure method, striking a balance between accuracy and computational efficiency [3]. KS-DFT is not a pure DFT, as it requires solving a self-consistent set of orbital equations [4]. In return, only a small fraction of the total energy, the exchange-correlation (XC) energy, need be approximated as a functional of the electronic spin densities. This produces far greater accuracy relative to a pure DFT such as Thomas-Fermi theory [5].

The computational bottleneck in KS-DFT calculations is the need to solve the KS equations, which formally scale as N^3, where N is the number of electrons. Thus there is strong interest in constructing an orbital-free DFT, avoiding this step [6]. A sufficiently accurate approximation to $T_{\rm S}[n]$, the kinetic energy (KE) of KS electrons, would produce an orbital-free scheme, greatly reducing the computational cost of DFT without sacrificing accuracy.

Many research efforts have recently focused in this direction [7]. Unfortunately, the relative accuracy requirements of a KE functional are much stricter than those of an XC functional, because the KE is typically comparable to the total energy of the system [5]. Worse, we also need accurate functional derivatives, since ultimately the density must be determined self-consistently via an Euler equation. The standard approximations using local and semi-local forms do not yield accurate derivatives [8].

Approximating $T_{\rm S}[n]$ has proven to be a difficult task for both extended and finite systems [6]. Some build on Thomas-Fermi theory with gradient expansions and various mixing coefficients [7] or on generalized gradient approximations (GGAs) with enhancement factors based on "conjointness" [9]. Others approximate $T_\theta = T_{\rm S} - T^{\rm W} \geq 0$, where $T^{\rm W}$ is the von Weizsäcker KE [10], or attempt to produce the correct linear response [6]. For multiple bonds, no present KE functional can accurately describe molecules far from equilibrium structures nor properly dissociate a diatomic. Moreover, solving the orbital-free Euler equation for GGA-like functionals can be difficult due to poor functional derivatives near nuclei [6].

A particularly difficult problem is to correctly dissociate a chemical bond. Any locally-based approximation has difficulties when the bond length is stretched to large distances. The fragments often contain fractional electron numbers, for which local approximations yield very wrong answers [11]. The worst case is the KE of a single bond in one dimension, where a local approximation yields fragments energies that are incorrect by a factor of 4 in the stretched limit.

To tackle these difficulties, we turn to machine learning (ML), a powerful tool for learning high-dimensional patterns via induction that has been very successful in many applications [12] including quantum chemistry [13–16]. Some of us recently suggested a new paradigm in density functional approximation using ML to approximate density functionals [17]. In that work, all particles were confined to a box so no densities looked like those

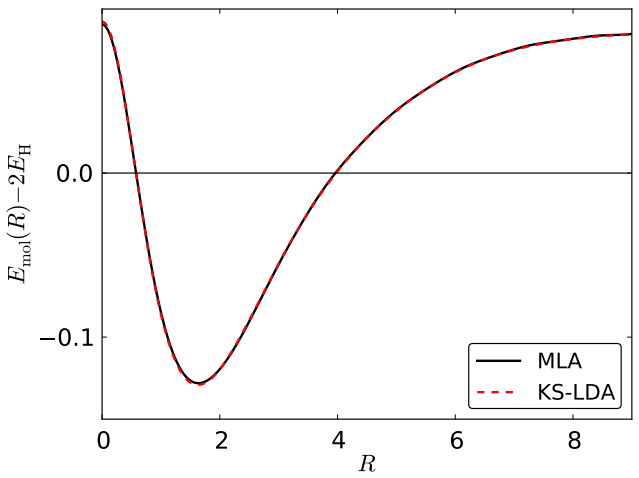

FIG. 1. The molecular binding curve for a 1d single bond (in atomic units). The machine learning approximation (MLA) curve is found found self-consistently, using 10 KS-LDA densities and kinetic energies (spaced evenly from $R = 0$ to 10) for training. Note that the incorrect large R limit is due to deficiencies in the LDA XC, *not* our MLA for the KE.

of well-separated atoms. In Fig. 1, we show what happens when we apply ML algorithms self-consistently to fermions in two wells as they are separated: ML has no particular difficulty in treating a situation that is virtually impossible with traditional DFT approximations. In this work, we construct an orbital-free KE functional based on ML that is capable of accurately describing 1d diatomics from the united atom limit to complete nuclear separation. Moreover, we obtain accurate self-consistent densities and molecular forces.

We use standard methods from ML [18], and atomic units throughout. Using kernel ridge regression (KRR), which is non-linear regression with regularization to prevent overfitting [18], our machine learning approximation (MLA) for the KE is

$$T^{\rm ML}(\boldsymbol{n}) = \sum_{j=1}^{M} \alpha_j k(\boldsymbol{n}_j, \boldsymbol{n}), \tag{1}$$

where α_j are weights to be determined, $\boldsymbol{n}_j$ are training densities, M is the number of training densities, and k is the kernel, which measures similarity between densities. We choose the Gaussian kernel

$$k(\boldsymbol{n}, \boldsymbol{n}') = \exp(-\|\boldsymbol{n} - \boldsymbol{n}'\|^2/(2\sigma^2)), \tag{2}$$

where σ is called the length scale. The weights are found by minimizing the cost function

$$\mathcal{C}(\boldsymbol{\alpha}) = \sum_{j=1}^{M} \Delta T_j^2 + \lambda \boldsymbol{\alpha}^T \boldsymbol{K} \boldsymbol{\alpha}, \tag{3}$$

where $\Delta T_j = T_j^{\rm ML} - T_j$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)$ and $\boldsymbol{K}$ is the kernel matrix, $\boldsymbol{K}_{ij} = k(\boldsymbol{n}_i, \boldsymbol{n}_j)$. The second term is a regularizer that penalizes large weights to prevent overfitting. Minimizing $\mathcal{C}(\boldsymbol{\alpha})$ gives

$$\boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \boldsymbol{T}, \tag{4}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{T} = (T_1, \ldots, T_M)$. The global parameters σ and λ are determined through leave-one-out cross-validation: For each density $\boldsymbol{n}'$ in the training set, the functional is trained on all densities except for $\boldsymbol{n}'$ and σ and λ are optimized by minimizing the absolute error on the $\boldsymbol{n}'$. Final values are chosen as the median over all optimum values. To test performance, the functional is always evaluated on new densities not in the training set.

Here, we consider a one-dimensional model of diatomic molecules, where the electron repulsion has the soft-Coulombic form [19]

$$v_{\rm ee}(u) = \frac{1}{\sqrt{1+u^2}}, \tag{5}$$

as this has been used in a variety of contexts. The one-body potential attraction of the two "nuclei" of nuclear charge Z at separation R is

$$v(x) = -Z(v_{\rm ee}(x - R/2) + v_{\rm ee}(x + R/2)), \tag{6}$$

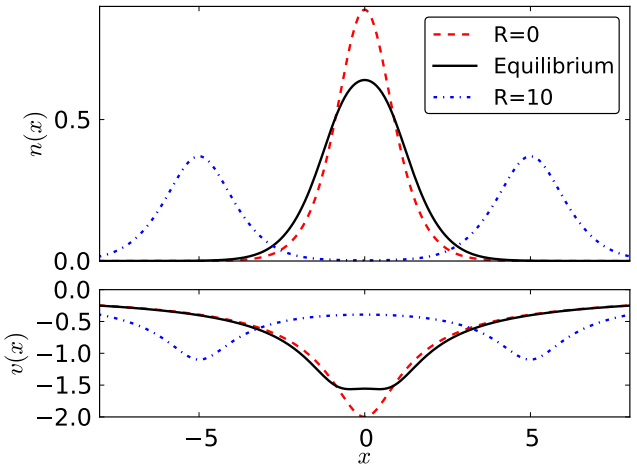

FIG. 2. The soft-Coulomb model of a diatomic for $Z = 1$, $N = 2$, which we test our ML method on. External potentials and corresponding KS densities are shown for $R = 0$ (dashed), equilibrium bond length at R=1.62 (solid), and stretched at $R = 10$ (dot-dashed), in atomic units.

and the internuclear repulsion is $V_{\rm NN}(R) = Z^2 v_{\rm ee}(R)$. We solve this model within KS-DFT [2] with the local density approximation (LDA) for XC [19, 20]. The spin-unpolarized form of LDA exchange for this system is given in Ref. [19], and an accurate parametrization of the LDA correlation energy is given in Ref. [20].

Our goal is to "learn" the non-interacting kinetic energy $T_{\rm S}[n]$ of the KS system. In previous work [17], where we demonstrated for the first time the ability of ML to approximate density functionals, the fermions were non-interacting and confined to live in a box, restraining the variety of possible densities. In particular, there was no analog of a binding energy curve where a density is centered on two sites whose separation varies continuously from small to infinite.

Fig. 2 shows the densities and potentials for the united atom, equilibrium bond length, and stretched diatomic. To generate a dissociation curve like that of Fig 1, we consider bond lengths up to $R = 10$, and place the entire system on a 500-point grid from $x = -20$ to 20. This is necessary to converge our KS-DFT calculations. We doubly-occupy the lowest Z orbitals, so that $N = 2Z$, where N is the number of electrons. We extract various energies and the density as a function of R for different values of N.

A curious point is that the ML method only needs 50 points to achieve the level of accuracy given in this paper *once the training energies are sufficiently accurate*. Thus, a 500-point calculation may be needed to find essentially exact energies, but a far smaller grid (50 points) then yields a sufficiently accurate representation of the densities for the ML to achieve arbitrarily accurate energies. ML automatically corrects for the coarseness of the grid.

To construct the model, we choose M training densities at evenly spaced R between 0 and 10 (inclu-

N	M	λ	σ	$\overline{\lvert\Delta T\rvert}$	$\lvert\Delta T\rvert^{\rm std}$	$\frac{\lvert\Delta T\rvert^{\rm max}}{\overline{\lvert\Delta T\rvert}}$
2	4	1.4e-3	105.	20	15	2.2
2	6	4.8e-3	5.2	2.9	3.6	4.1
2	8	4.1e-5	39.3	2.6	2.8	3.7
2	10	3.0e-7	32.5	0.13[0.93]	0.19[2.6]	6.1[14]
2	12	5.6e-7	15.9	0.092	0.14	6.1
2	15	2.4e-8	17.8	0.045	0.047	3.9
2	20	2.4e-8	15.2	0.038	0.037	3.5
2	25	1.9e-10	6.1	0.002	0.002	3.9
4	10	3.3e-7	87.7	0.25[1.3]	0.31[3.7]	4.9[15]
4	20	1.1e-10	46.1	0.02	0.018	3.9
6	10	8.1e-6	74.5	3.1[13]	4.0[33]	4.7[11]
6	20	1.8e-9	55.1	0.016	0.018	4.8
8	10	2.0e-6	56.2	1.4[7.7]	1.7[22]	5.0[14]
8	20	2.9e-9	16.1	0.016	0.012	3.3

TABLE I. Parameters and errors (in kcal/mol) as a function of fermion number N and number of training densities M. Mean, standard deviation and max values taken over 199 test densities evenly spaced from $R = 0$ to 10 (inclusive). Brackets represent errors on self-consistent densities.

sive). Table I shows the performance of the MLA, evaluated on a test set with $M = 199$. To compare, we tested the LDA in 1d, $T^{\rm loc}[n] = \pi^2 \int dx\, n^3(x)/6$, and a modified gradient expansion approximation (GEA) [21], $T^{\rm MGEA}[n] = T^{\rm loc}[n] + cT^{\rm W}[n]$, where $c = 0.559$ has been chosen to minimize the error. LDA has a MAE of 45 kcal/mol, while the GEA only improves that to 33 kcal/mol. For $N = 2$, we have already achieved a MAE below 1 kcal/mol for $M = 10$. No other existing approximations achieve this level of accuracy, which can be systematically improved.

Thus far, our results have been reported evaluating the ML functional on exact densities. In real applications of orbital-free DFT, these are not available. The density that minimizes the approximate energy satisfies:

$$\frac{\delta T_{\rm S}[n]}{\delta n(x)} = \mu - v_{\rm S}(x), \tag{7}$$

where μ is a constant. But just as in Ref. [17], ML does *not* yield an accurate functional derivative. Fig. 3 shows that the derivative of our MLA evaluated at the ground-state density is very different from the exact one. The exact functional derivative shows how the KE changes along *any* direction, but the model only knows about directions along which the training densities lie. Since these densities are generated from a set of potentials parametrized by R, the data is effectively 1d (locally), embedded in a high-dimensional space. The same effect occurs in [17], except there we had 9 parameters in the potential.

The bare gradient of the MLA makes solving Eq. 7 unstable. If we project the gradient onto the space of the data, we can extract the projected functional derivative of the MLA (see Fig. 3). In [17], we used principal component analysis (PCA) to locally approximate the neighborhood of training densities as linear, but here the local neighborhood is nonlinear, and the PCA method leads to inaccurate self-consistent densities. Instead, we use the nonlinear gradient de-noising (NLGD) projection method developed in Ref. [23].

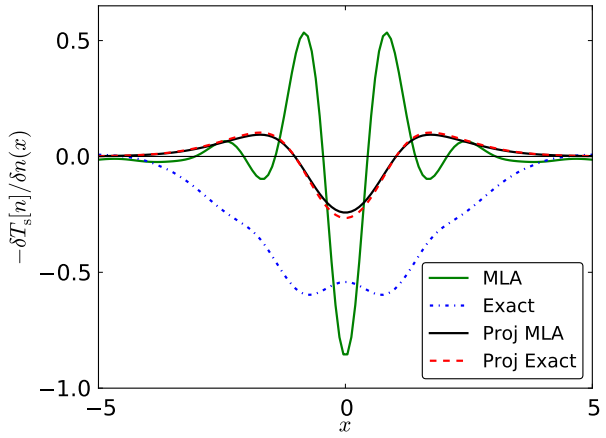

FIG. 3. The functional derivative of our MLA (green) cannot reproduce the exact derivative (blue dot dashed, given by $-v_{\rm S}[n]$ if evaluated at the ground-state density) because this information is not contained in the data. However, both agree when projected onto the space of the data (black and red dashed). Shown for $Z = 1$ and $M = 10$ at equilibrium bond length (R=1.62), in atomic units.

Applying the NLGD projected gradient descent technique [23] to minimize Eq. 7 with $T^{\rm ML}$ yields the results in Fig. 4, where we plot the error in total energy, $\Delta E = E^{\rm ML}[\tilde{n}] - E[n]$, where n is the exact density and $\tilde{n}$

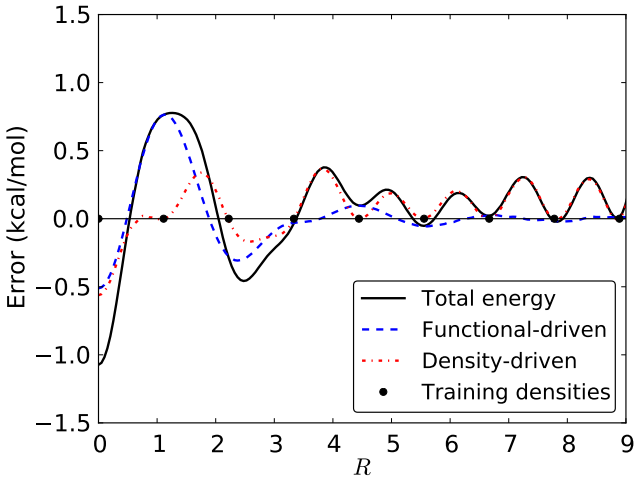

FIG. 4. The total error of the model and the functional- and density-driven errors [22]. The dots mark the location of the training densities. R given in atomic units, for $Z = 1$ and $M = 10$.

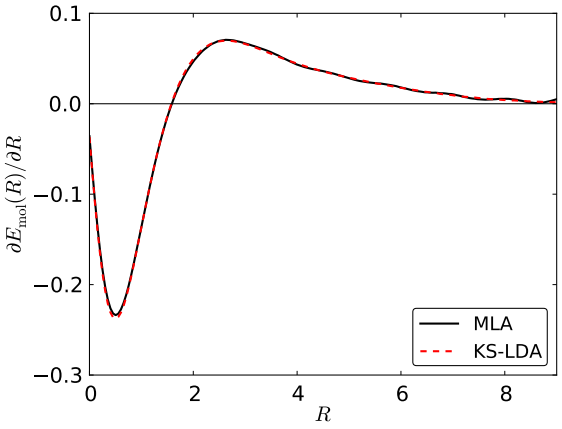

FIG. 5. Molecular forces as a function of R, for $Z = 1$ and 10 training densities. Derivatives are calculated via finite-difference.

is the self-consistent density. This is split into the error due to the approximate functional (i.e. functional-driven error), $\Delta E_{\rm F} = E^{\rm ML}[n] - E^{\rm LDA}[n]$, and the error due to the deviation of the self-consistent density from the exact (i.e. density-driven error), $\Delta E_{\rm D} = E^{\rm ML}[\tilde{n}] - E^{\rm ML}[n]$ [22]. Near $R = 0$, $\Delta E_{\rm F}$ is larger because the $T_{\rm s}$ is rapidly changing. $\Delta E_{\rm D}$ is zero at the training densities and is largest farthest from them. At large R, the $\Delta E_{\rm D}$ dominates over $\Delta E_{\rm F}$. Table I gives the MAE of the ML functional evaluated on the self-consistent densities for different N and M.

Fig. 5 shows the forces calculated with model densities. The forces are very accurate and should be suitable for, e.g., an *ab-initio* molecular dynamics calculation. Combined with the reduced computational cost of orbital-free DFT, this method has the potential to simulate very large systems at the same level of accuracy KS-DFT currently provides.

Our work shows that, for a one-dimensional system, machine learning can be used to find a density functional that produces accurate energies and forces on self-consistent densities, even when bonds break. No existing orbital-free scheme comes close to this level of accuracy. Although this example is limited to one dimension, there is no reason in principle to doubt the efficacy of the method for real bonds.

The authors thank IPAM at UCLA for hospitality and acknowledge NSF Grant No. CHE-1240252 (JS, KB), EU PASCAL2 and DFG Grant No. MU 987/4-2 (K.H., and K. R. M.), EU Marie Curie Grant No. IEF 273039 (M. R.), and NRF Korea Grant No. R31-10008 (K. R. M.).

[1] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. B **136**, 864–871 (1964).
[2] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. A **140**, 1133–1138 (1965).
[3] Kieron Burke, "Perspective on density functional theory," J. Chem. Phys. **136**, 150901 (2012).
[4] Kieron Burke and Lucas O. Wagner, "Dft in a nutshell," Int. J. Quant. Chem **113**, 96–101 (2013).
[5] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem* (Springer, 1990).
[6] V.V. Karasiev and S.B. Trickey, "Issues and challenges in orbital-free density functional calculations," Comput. Phys. Commun. **183**, 2519 – 2527 (2012).
[7] Valentin V. Karasiev, Randy S. Jones, Samuel B. Trickey, and Frank E. Harris, "Recent advances in developing orbital-free kinetic energy functionals," in *New Developments in Quantum Chemistry*, edited by José Luis Paz and Antonio J. Hernández (Transworld Research Network, Kerala, India, 2009) pp. 25–54.
[8] C. J. Umrigar and Xavier Gonze, "Accurate exchange-correlation potentials and total-energy components for the helium isoelectronic series," Phys. Rev. A **50**, 3827–3837 (1994).
[9] V. V. Karasiev, R. S. Jones, S. B. Trickey, and Frank E. Harris, "Properties of constraint-based single-point approximate kinetic energy functionals," Phys. Rev. B **80**, 245120 (2009).
[10] Carl Friedrich von Weizsäcker, "Zur Theorie der Kernmassen," **96**, 431–458 (1935).
[11] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory," Science **321**, 792–794 (2008).
[12] Klaus-Robert Müller, Sebastian Mika, Gunnar Rätsch, Koji Tsuda, and Bernhard Schölkopf, "An introduction to kernel-based learning algorithms," IEEE Trans. Neural Netw. **12**, 181–201 (2001).
[13] O. Ivanciuc, "Applications of support vector machines in chemistry," in *Reviews in Computational Chemistry*, Vol. 23, edited by Kenny Lipkowitz and Tom Cundari (Wiley, Hoboken, 2007) Chap. 6, pp. 291–400.
[14] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," Phys. Rev. Lett. **104**, 136403 (2010).
[15] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," Phys. Rev. Lett. **108**, 058301 (2012).
[16] Zachary D. Pozun, Katja Hansen, Daniel Sheppard, Matthias Rupp, Klaus-Robert Müller, and Graeme Henkelman, "Optimizing transition states via kernel-based machine learning," J. Chem. Phys. **136**, 174101 (2012).
[17] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke, "Finding density functionals with machine learning," Phys. Rev. Lett. **108**, 253002 (2012).
[18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman,

The Elements of Statistical Learning. Data Mining, Inference, and Prediction, 2nd ed. (Springer, New York, 2009).

[19] Lucas O. Wagner, E. M. Stoudenmire, Kieron Burke, and Steven R. White, "Reference electronic structure calculations in one dimension," Phys. Chem. Chem. Phys. **14**, 8581–8590 (2012).

[20] N. Helbig, J. I. Fuks, M. Casula, M. J. Verstraete, M. A. L. Marques, I. V. Tokatly, and A. Rubio, "Density functional theory beyond the linear regime: Validating an adiabatic local density approximation," Phys. Rev. A **83**, 032503 (2011).

[21] Donghyung Lee, Lucian A. Constantin, John P. Perdew, and Kieron Burke, "Condition on the kohn–sham kinetic energy and modern parametrization of the thomas–fermi density," J. Chem. Phys. **130**, 034107 (2009).

[22] Min-Cheol Kim, Eunji Sim, and Kieron Burke, "Classifying and reducing errors in density functional calculations," arXiv:1212.3054 (2013).

[23] John C. Snyder, Matthias Rupp, Katja Hansen, Klaus-Robert Müller, and Kieron Burke, "Accurate densities from inaccurate functional derivatives," In Prep. (2013).