Direct calculation of the kinetic energy functional derivative using Machine Learning

HBM. Saidoune1,∗ S. Kais2, and FH. Alharbi3,4
1 Qatar Environment and Energy Research Institute (QEERI), Hamad Bin Khalifa University, Doha, Qatar
2 Department of Chemistry, Physics, and Birck Nanotechnology Center, Purdue University
3 Electrical Engineering Department, KFUPM, Dharan, Saudi Arabia
4 K.A.CARE Energy Research & Innovation Center, Dharan, Saudi Arabia

We report a direct scheme calculation of kinetic energy functional derivative using Machine Learning. Support Vector Regression and Kernel Ridge Regression techniques were independently employed to estimate the kinetic energy functional and its derivative. Even though the accuracy should have been a decisive factor in modeling a realistic functional, we show that at a certain level it affects the generalizability of the model. By choosing the right regularization term and by considering a reasonable interplay between it and the accuracy, we were able to deduce the functional derivative from a model that was trained to estimate the kinetic energy. Although the derivative calculations demand very high accuracy to account for small variations of the kinetic energy, the developed estimator was capable of capturing these extremely small changes of the electron density. This work pours into highly effective implementation of the orbital-free density functional theory as it employs only direct calculation schemes.

Atomic scale calculation has become one of the essential methodologies in nowadays scientific activities. Amongst a large number of its methods and techniques, density functional theory (DFT) appears to be the most popular and versatile quantum mechanical theory in investigating the electronic structure. It was introduced in the seminal theory of Thomas and Fermi1,2, had its theoretical foundations laid by Hohenberg and Kohn3 and made practical by adopting the Kohn-Sham fictitious system scheme4. The Kohn-Sham DFT has witnessed an unprecedented success story in showcasing both the accuracy and the low computational cost. It was successful in describing many material properties and - to a good extent - was an efficient tool to investigate chemical systems. However, DFT suffers when it comes to modeling bulk solids band gaps, Van der Waals interactions and strongly correlated systems5–8 in addition to the high computational cost associated with nowadays demanding scientific problems. In spite of being a direct manifestation of the density-based Hohenberg-Kohn theorems, DFT scheme uses Kohn Sham orbitals in order to calculate densities, rendering the computation prohibitively expensive when investigating systems with large number of electrons N. Meanwhile many attempts have been made in order to tackle these flaws, those trials resulted in ameliorating the overall accuracy by improving the pseudo-potentials and the exchange-correlation (XC) functionals. However much less improvement has been achieved in lowering the computational cost9.

To this end, orbital-free density functional theory10 (OFDFT) (the original Kohn-Hohenberg theory) seemed to be a good alternative to the current stagnated orbital-based density functional theory from calculation cost point of view. Nevertheless, its practicality was every time faced by the long-standing issue of its functionals accuracy. Kinetic energy is the leading term of an electronic system total energy, therefore, errors made in approximating it have a dramatic impact on the total energy accuracy. Even though, considerable improvement has been made in calculating and approximating exchange-correlation functionals4,11–15, the kinetic (Kohn Sahm) energy remains a bottleneck to tackle for a full DFT-orbital free implementation. The first analytical expressions of the kinetic energy functionals was given by the Thomas Femi functional

$$T_{TF}[n(r)] = c_D \int n(r) \frac{Dn^2}{D^2 r} d^D r$$

for uniform densities and by the Von Weizsacker functional

$$T_{vW}[n(r)] = \frac{1}{8} \int \frac{\|\nabla(n(r))\|^2}{n(r)} d^D r$$

for single orbital systems, respectively. Both functionals are given in a $D$ dimensional space where $c_D$ is a $D$-dependent constant. Following that, many attempts have been devoted to get an accurate KE functional, ranging from proposing linear combination of the aforementioned functionals to employing conventional gradient expansion and enforcing linear response behavior, these attempts were shiny for some systems but were overall non-transferable.17–25

Recently, new methods which are based on learning from data, have been proposed to approach the OFDFT30–33 from a different perspective. Although an overall good accuracy has been achieved throughout these attempts, the applicability of some of these methods are limited when coming to the functional derivative calculations.

In this letter, we develop a method based on Machine Learning (ML) that outputs accurate kinetic energy functionals, we then calculate the functional derivative, in a direct manner, without any further training. We use support vector regression (SVR)34 and kernel ridge regression (KRR) tools throughout this work. We will give an overview on SVR and refer the reader to Ref. [30] for a detailed implementation of KRR in predicting KE functionals. SVR is mainly based on the kernel mapping
new objective function as the following:

\[
\text{minimize} \left( \frac{1}{2} \| \Theta \|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \right)
\]

subject to \[
\begin{align*}
        y_i - (\Theta^T x_i + b) & \leq \epsilon + \xi_i \\
(\Theta^T x_i + b) - y_i & \leq \epsilon + \xi_i^* \\
\xi_i, \xi_i^* & \geq 0,
\end{align*}
\]

where the slack variables \(\xi_i\) and \(\xi_i^*\) are introduced to account for model errors. By introducing Lagrange multipliers to account for the constraints one ends up with intuitive rather simple expressions of the estimator function and the weights vector, namely:

\[
f(x, \Theta) = \sum_{i=1}^{n_{sv}} (\alpha_i - \alpha_i^*)(x_i, x)
\]

\[
\Theta = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*)x_i
\]

We refer the reader to Ref (SVR) for a detailed passage from Eq. 6 to Eq. 7 and 7. The dual variables \(\alpha_i\) and \(\alpha_i^*\) are Lagrange multipliers introduced to account for the first and second constraints in Eq. 6. \(f(x, \Theta)\) is the regression function and is given as a linear combination of inner products between the input variable and other training samples. This very simple and sophisticated formula can be extended to non-linear systems by adopting the kernel trick, whereby the original input \(x\) is mapped into a vector \(\Psi(x)\) from the feature space (a higher dimensional space). In the new feature space, the dimensionality expansion plays an essential role in rendering the problem linear. Throughout this paper, we use the Gaussian kernel,

\[
G(x, \gamma) = \exp \left( -\frac{\|x - x\|^2}{2\gamma^2} \right),
\]

where \(\gamma\) is the variance of the distribution. An illustration of the role of \(\gamma\) and other hyperparameters is given in the supplemental material. Similarly to Eq. 7, the feature space estimator can be expressed as function of the mapped vectors \(\Psi(x)\), thus reads:

\[
\Xi(\Theta, x) = \sum_{i=1}^{n_{sv}} (\alpha_i - \alpha_i^*)K(x_i, x).
\]

\(K(x_i, x)\) is called the kernel and is given by the inner product between mapped vectors, namely \(K(x_i, x) = \langle \Psi(x_i), \Psi(x) \rangle\). Moreover, we keep the same expression for the weights vector \(\Theta\) whereby we substitute the input vectors by the new feature space mapped vectors, \(\Psi(x)\):

\[
\Theta = \sum_{i=1}^{m} (\alpha_i - \alpha_i^*)\Psi(x_i).
\]

The impact of the kernel types as well as the model hyperparameters are showcased for illustrative examples in the supplementary material.
In Fig. 1, we used the estimator of Eq. (10) in order to calculate the kinetic energy as function of the densities principal components (c) and kernel-principal components (d). We used two fold cross validation with linear regression for training purpose. As expected, the increase of the number of PCs (or K-PCs) reduced the mean absolute error (MAE) in the KE calculation. While the error calculation in (c) saturated to a value 0.27 HA, the error in (d) headed to a very tiny range, an evidence of a readily non-linear problem. Insets of (c) and (d) are 3 dimensional visualization of the MAE as function of the first two principal components (containing most of the variance) without and with applying the kernel, respectively. Inset of (c) shows a compelling 3 dimensional behavior while a tendency of dimensionality reduction can be felt in inset of (d). In (c) MAE reached a plateau of constant error after including 7 PCs. This number increased after applying the kernel to 20 K-PCs. The application of the kernel was associated with a mapping to a higher dimensional space, thus explaining the need of larger number of K-PCs to account for the full variance.

During the training, the minimization of the loss function (see Eq. 4) leads to the obtention of the Lagrange multipliers \( \alpha \) \( i \) and \( \alpha^* \). These parameters are used to estimate the kinetic energy according to Eq. 7. Saying that, one needs to choose carefully a different category of parameters which are not included in the optimization of the loss function, are instead user defined. These hyperparameters can be chosen by cross-validating the training data and sorting the corresponding MAE. Figure 2, depicts a contour plots of the KE estimation scores as function of \( \gamma \) and \( C \) for the SVR model, and of \( \gamma \) and \( \alpha \) for the KRR model. In both plots \( \gamma \) is a measure of the kernel distribution variance. \( C \) is a regularization term that penalizes the model error in SVR and \( \alpha \) is a regularization term that weights the L2 norm term in KRR. A decimal logarithmic scale is applied in both plots. In Fig. 2(c), an increase of \( C \) favors higher scores (less estimation errors on the cross-validation set) which reached 0.9 for \( C > 10^{-1} \). The optimal values of \( \gamma \) spans the range \( 10^{-5} \) to \( 10^{-2} \). In a similar fashion, the optimal range of the KRR hyperparameters coincides with \( \gamma \) spans the interval of values less than 1. A small value of \( \alpha \) leads to a better estimation score as \( \alpha \) plays the role of the inverse of the regularizer term in other regression based ML algorithms (like 1/2\( C \) in the case of logistic regression). Supported by the data of Fig. 2, we choose \( \epsilon \) to be equal to \( 5 \times 10^{-4} \), \( C = 100 \) and \( \gamma = 5 \times 10^{-3} \) for SVR, \( \alpha \) and \( \gamma \) take on the values \( 10^{-5} \) and 0.07, respectively for the KRR. In this letter, as we mention the importance of choosing the convenient hyperparameters that could result in a good accuracy of the KE calculation, we emphasize, on equal footing, the importance of a regularized algorithm that can be generalized for the KE derivative calculation. In table 1. (see supplemental material) we report the MAE calculated on the test set while evaluating the KE. Obviously MAE decreases when we increase the dimension of the training set, it does not however follow a monotonic behavior when \( \epsilon \) decreases. In addition to that, we define a mean absolute error for the KE derivative, which we denote DMAE (for derivative mean absolute error). In the derivative case, an average of the errors over the real space mesh defining the derivative is calculated first,

\[
E_1 = \frac{1}{n_x} \sum_{i=1}^{n_x} \left| \frac{\delta \xi(\theta, x)}{\delta n(x)} - \frac{\delta T}{\delta n(x)} \right|. \tag{12}
\]

An averaging over the test set is needed to insure the generalization of the derivative error behavior.

\[
DMAE = \frac{1}{t} \sum_{j=1}^{t} E_j \tag{13}
\]

In contrast to MAE, the derivative error does not exhibit a monotonic behavior as function of the accuracy hyperparameters \( \epsilon \) (SVR) and \( \alpha \) (KRR) (see Fig. 1 in suplemental material). Both plots witnessed an increasing accuracy when \( \epsilon \) (\( \alpha \)) increased up to a certain value \( \epsilon_c \). For \( \epsilon > \epsilon_c \), the error in the derivative reaches higher values despite the good accuracy of the KE functional. The key feature in this non usual behavior is known as the regularization. Indeed, the calculation of the KE derivative demands the evaluation of the KE and its variation \( n_x \) times (\( n_x \) being the number of mesh points), hence the KE estimator had to be generalizable in order to account for these different evaluations. As we have seen in the preceding text, \( \epsilon \) and \( \alpha \) are accuracy terms but their increase could lead to overfitting the data. When calculating the derivatives, the KE estimator needs to be evaluated at small variations of the density. The overfitted model will not be capable of predicting the small variation contribution to the KE estimator.

In this work, we consider the cases of potentials consisting of three different Gaussian dips between infinite walls such that

\[
v(x) = \begin{cases} 
-\sum_{i=1}^{3} a_i \exp \left( \frac{-(x-b_i)^2}{2c_i^2} \right) & 0 < x < L \\
\infty & \text{otherwise}
\end{cases}
\]  

\( \text{where } a_i, b_i \text{ and } c_i \text{ are random distributions ranging from } \{1-10\}, \{0.2-0.8\} \text{ and } \{0.03-0.1\} \text{ respectively and } L \text{ is the width between the two infinite walls. This potential was used for similar machine learning calculations}^{30,33} \text{. We use the Galerkin method}^{35} \text{ for the seek of the exact solution. The solution to the considered potential is an array of densities projected on the grid of the confining space which we save in the array } \mathbf{x}, \text{ the corresponding kinetic energies are stored in the vector } y.

The given results of Fig. 2 predict a robust estimator of the kinetic energy as it has been shown to perform well with high accuracies (Fig. 2(a)) and low variances (Fig. 2(b)) on the cross-validation and test sets. The accuracies reported in the current work are comparable to what has been recently reported by groups using similar techniques\(^{30}\). Nevertheless, we think that the biggest
As such, the functional derivative can be expressed in the following way:

\[
\lim_{\eta \to 0} \frac{F[f + \eta \phi] - F[f]}{\eta} = \frac{d}{d\eta} F[f + \eta \phi] \bigg|_{\eta=0},
\]

where \( F \) is the functional to be derived, \( \eta \) is a an infinitesimal coefficient and \( \phi \) is a random function. We use the left hand side of Eq. 17 as a numerical implementation of the mathematical formula. A delta distribution \( \delta \) is used instead of \( \phi \) to account for the variation of the kinetic energy functional, such a choice is seen to be mathematically correct but also handy for our numerical implementation.

Fig. 2 showcases the external potential calculated from machine learning by differentiating the kinetic energy estimator the way dictated by Eq. 17. The real potentials are present for reference. Each three consecutive plots (from left to right) designate same number of dips of the original potential with different random magnitude, center and spread (a, b and c in Eq. 14). As has been mentioned in the preceding text, the functional differentiation is done without any additional training. Instead, the KE estimator is differentiated as in Eq. 17. Although the accuracy of the original KE estimator is not enough to account for the infinitesimal parameter \( \eta \), the inclusion of a proper standardization of the input data reduces the discrepancy between \( f \) and \( \eta \phi \) (in Eq. 17) thus, makes the evaluation of \( f + \eta \phi \) falling within the range of accuracies tolerated by the estimator. Here we choose \( \eta \) to be equal to \( 10^{-9} \). It is worth noting that in some cases, larger discrepancies might be detected between the ML calculated potentials and the real ones. These mostly happen at flat surfaces of the potential and might be caused by inevitable numerical errors while evaluating the functionals derivatives. The first three rows of Fig. 3 are results obtained using the KRR algorithm while the remaining row showcases the derivatives obtained using SVR for training the kinetic energy functional. The training data set used to train the KE estimator was a mixture of potentials with random parameters and with different number of dips. We believe that this choice was an enhancing factor which led to more generalizable KE estimator which has same derivative characteristic as the real one. Looking at the good match between real and ML calculated potentials, one could arguably think the good match is caused by the highly accurate KE estimator which could cope with small variation of the electron density. However, (see Fig. 1 in supplemental material) the accuracy is a double-edged parameter that can also promote overfitting. The good choice of the standardization of the input in addition to the choice of more general input data play a role as important as the accuracy of the estimator in giving correct functional derivatives.

In the current work, we have reported highly accurate kinetic energy functional estimator based on ma-
We have shown that it is possible to obtain, in a direct way, the functional derivative of the kinetic energy estimator without any further training. Throughout this letter, we emphasized the importance of a careful choice of the hyperparameters in addition to choosing a well studied standardization technique. The model we present can be further ameliorated by considering datasets for more general physical systems or/and by applying numerical techniques that can account for physical constraints.

* hsaidhouni@hbku.edu.qa

1 L. H. Thomas, Math. Proc. Cambridge Philos. Soc. 23, 542 (1926).
2 E. Fermi, Zeits. f. Physik 48, 73 (1928).
3 P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
4 W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
5 J. Klime and A. Michaelides, J. Chem. Phys. 137, 120901 (2012).
6 K. Burke, J. Chem. Phys. 136, 150901 (2012).
7 A. D. Becke, J. Chem. Phys. 140, 18A301 (2014).
8 R. O. Jones, Rev. Mod. Phys. 87, 897 (2015).
9 A. Pribram-Jones, D. A. Cross and K. Burke, Annual Review of Physical Chemistry, 66, 283 (2015).
10 V. L. Liguori and E. A. Carter, (2005) An Introduction to Orbital-Free Density Functional Theory. In Handbook of Materials Modeling. Springer, Dordrecht.
11 P. Perdew, Phys. Rev. B 33, 8822 (1986).
12 A. D. Becke, Phys. Rev. A 38, 3098 (1988).
13 C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B 37, 785 (1988).
14 A. D. Becke, J. Chem. Phys. 98, 5648 (1993).
15 J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
16 C. F. Von Weizsacker, Z. Phys. 96, 431 (1935).
17 P. K. Acharya, L. J. Bartolotti, S. B. Sears, and R. G. Parr, Proc. Natl. Acad. Sci. 77, 6978 (1980).
18 P. Garcia-Gonzalez, J. E. Alvarado, and E. Chacon, Phys. Rev. A 54, 1897 (1996).
19 D. A. Kirzhnits, Sov. Phys. JETP, 5, 64 (1957).
20 C. H. Hodges, Can. J. Phys. 51, 1428 (1973).
21 L. W. Wang and M. P. Teter, Phys. Rev. B 45, 13196 (1992).
22 E. Smargiassi, P. A. Madden, Phys. Rev. B 49, 5220 (1994).
23 M. Foley and P. A. Madden, Phys. Rev. B 53, 10589 (1996).
24 T. Gá and A. Nagy, J. Mol. Struct. 501, (2000).
25 E. Sim, J. Larkin, and K. Burke, J. Chem. Phys. 118, 8140 (2003).
26 T. A. Wesolowski and W. A. Wang (2013) Recent progress in orbital-free density functional theory. In Recent Advances in Computational Chemistry. World Scientific, Hoboken, New Jersey.
27 M. Levy, Proc. Natl. Acad. Sci. (USA) 76, 6062 (1979).
28. M. Levy, Bull. Amer. Phys. Soc. **24**, 626 (1979).
29. M. Levy, Phys. Rev. A **26**, 1200 (1982).
30. J. C. Snyder, M. Rupp, K. Hansen, K-R. Miller, and K. Burke, Phys. Rev. Lett. **108**, 253002 (2012)
31. J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K-R. Miller, and K. Burke, J. Chem. Phys. **139**, 224104 (2013)
32. F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke and K-R. Miller, Nature Communicationsvolume **8**, 872 (2017)
33. F. H. Alharbi and S. Kais, Int J Quantum Chem. **117**, e25373 (2017)
34. C. Cortes and V. Vapnik, Mach Learn, **20**, 273 (1995)
35. A. Ern, J.L. Guermond, Theory and practice of finite elements, Springer, (2004)
36. R. G. Parr, Density-Functional Theory of Atoms and Molecules In: Fukui K., Pullman B. (eds) Horizons of Quantum Chemistry, vol 3. Springer, Dordrecht