Learning Green's Functions in RKHS and Application to Kohn–Sham DFT

Haejung Koh, Pritam Kayal

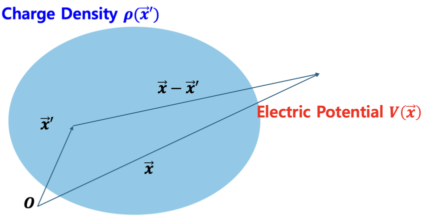
16th April 2025

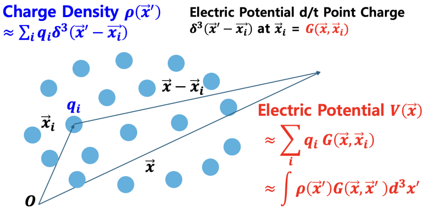
Based on: Stepaniants et al., "Learning Partial Differential Equations in Reproducing Kernel Hilbert Spaces," JMLR, 2023

Clarification of the Paper

- Paper: Stepaniants et al., "Learning Partial Differential Equations in Reproducing Kernel Hilbert Spaces," JMLR, 2023
- Goal: Learn the Green's function G(x,y) of linear PDEs via functional regression in RKHS
- Key idea: Represent G(x, y) as a kernel expansion, minimize operator residual norm in RKHS dual space
- Target PDE form: $u(y) = \int G(x, y) f(x)\, dx + \beta(y)$

Electric Potential from Point Charge Sources





Gauss Law: $-\vec{\nabla}^2 V(\vec{x}) = \rho(\vec{x})$

$-\vec{\nabla}^2$: $-$ Laplacian

$V(\vec{x})$: Electric Potential

$\rho(\vec{x})$: Charge Density

Electric potential due to point charge is given by $G(\vec{x},\vec{x}') = 1/4\pi|\vec{x}-\vec{x}'|$. Full potential is a sum of them:

$$V(\vec{x}) = \int d^3x' \frac{\rho(\vec{x}')}{4\pi|\vec{x}-\vec{x}'|}.$$

Heuristic Thinking on Green's Function

Green's Function $G(x, s)$ is a solution of a linear PDE for point source at s given by Dirac Delta $\delta(x-s)$:

$$\hat{P}_x G(x, s) = \delta(x - s).$$

For $\hat{P}_x u(x) = f(x)$ and $f(x) \approx \sum_i c_i \delta(x - s_i)$ (RKHS-like), we have

$$\hat{P}_x[\sum_i c_i G(x, s_i)] = \sum_i c_i[\hat{P}_x G(x, s_i)] = \sum_i c_i \delta(x - s_i) \approx f(x). \tag{1}$$

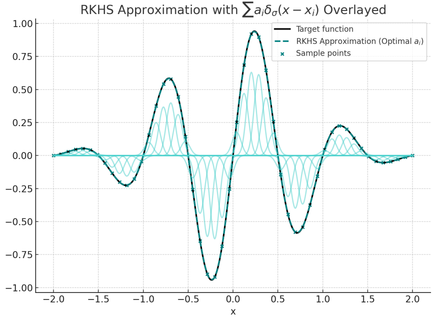
Then u is given by a linear superposition of G from different sources s_i:

$$u(x) \approx \sum_i c_i G(x, s_i) = \sum_i c_i \int \delta(x' - s_i) G(x, x') dx' \tag{2}$$

$$= \int [\sum_i c_i \delta(x' - s_i)] G(x, x') dx' \approx \int f(x') G(x, x') dx'. \tag{3}$$

Heuristic Thinking on Green's Function

In practice, we can find a "small enough" σ where $k_\sigma(x, x') \equiv \delta_\sigma(x - x')$ serve as a "pseudo" Dirac Delta.



$$f(x) = \sin(2\pi x)e^{-x^2} \approx \sum_{i=1}^{n} a_i \delta_\sigma(x - x_i)$$

$$\sigma = 0.05, \quad n = 50$$

- In RKHS, we have "Dirac Delta-like" reproducing kernel $k(\cdot, \cdot)$.
- Green function would be a solution of $\hat{P}_x G(x, y) = k(x, y)$.
- RKHS would be a "right" function space to solve PDE.

Setting up our working space

Definition

For any compact $K \subset \mathbb{R}^n$, $\mathcal{D}_K$ is defined as the set of smooth functions ϕ with $\text{supp}(\phi) \subset K$ with a topology τ_K defined by a countable collection of norms

$$||\phi||_N = \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^n} |D^\alpha \phi(x)|$$

This makes $\mathcal{D}_K$ a Frechet Space (a complete topological vector space with a translation invariant metric)

Definition

$\mathcal{D}(\mathbb{R}^n)$ is defined as the set of smooth functions with compact support. We define $\beta = \{W \subset \mathcal{D}(\mathbb{R}^n) : W \text{ is convex and balanced}, W \cap \mathcal{D}_K \in \tau_K \text{ for all compact } K \subset \mathbb{R}^n\}$. We equip $\mathcal{D}(\mathbb{R}^n)$ with topology

$$\tau = \{\phi + W : \phi \in \mathcal{D}(\mathbb{R}^n), W \in \beta\}$$

Setting up our working space (cont.)

Definition

A continuous (w.r.t τ) linear functional on $\mathcal{D}(\mathbb{R}^n)$ is called a distribution on $\mathbb{R}^n$. The set of distributions is denoted $\mathcal{D}'(\mathbb{R}^n)$.

Examples:

- Dirac distribution δ is defined as $\langle \delta, \phi \rangle = \phi(0)$
- If f is any locally integrable function, that is $\int_K |f(x)|dx < \infty$ for all compact $K \subset \mathbb{R}^n$, then Λ_f defined as $\langle \Lambda_f, \phi \rangle = \int f(x)\phi(x)dx$ is a distribution
- If Λ is any distribution then for any multi-index $\alpha \in \mathbb{R}^n$, then $D^\alpha \Lambda$, defined as $\langle D^\alpha \Lambda, \phi \rangle = (-1)^{|\alpha|} \langle \lambda, D^\alpha \phi \rangle$, is also a distribution. Here

$$D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \left(\frac{\partial}{\partial x_2}\right)^{\alpha_2} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}$$

Convolution

For two function f and g on $\mathbb{R}^n$, we define their convolution as

$$(f * g)(x) = \int f(y)g(x-y)dy \text{ (when it exists)}$$

For a distribution $\Lambda \in \mathcal{D}'(\mathbb{R}^n)$ and $\phi \in \mathcal{D}(\mathbb{R}^n)$, $\Lambda * \phi$ is a function on $\mathbb{R}^n$ defined as

$$(\Lambda * \phi)(x) = \langle \Lambda, \tau_x \tilde{\phi} \rangle \text{ where } \tau_x f : y \mapsto f(y-x) \text{ and } \tilde{f} : y \mapsto f(-y)$$

Properties:

- $\delta * \phi = \phi$ for any $\phi \in \mathcal{D}(\mathbb{R}^n)$
- $\Lambda * \phi \in C^\infty(\mathbb{R}^n)$ for any $\Lambda \in \mathcal{D}'(\mathbb{R}^n), \phi \in \mathcal{D}(\mathbb{R}^n)$ and

$$D^\alpha(\Lambda * \phi) = (D^\alpha \Lambda) * \phi = \Lambda * (D^\alpha \phi)$$

Technique to Find Solutions to Linear PDEs

Definition

If P is a polynomial function on $\mathbb{R}^n$ defined as $P(x) = \sum_{|\alpha| \leq N} c_\alpha x^\alpha$ then we define differential operator $P(D)$ as $P(D) = \sum_{|\alpha| \leq N} c_\alpha D^\alpha$. Here, x^α for $x \in \mathbb{R}^n$ and multi-index $\alpha \in \mathbb{R}^n$, is defined as $\prod_{j=1}^n x_i^{\alpha_i}$

Theorem

For any non-homogeneous linear PDE, $P(D)u = f$ for some polynomial P, if there exists a distribution Λ (called a fundamental solution) such that $P(D)\Lambda = \delta$ then $u = \Lambda * f$ is a solution.

Proof.

Let Λ be such a distribution. Then, using the properties above,

$$P(D)(\Lambda * f) = \big(P(D)\Lambda\big) * f = \delta * f = f$$

$\square$

Green's Functions

Note that if Λ_F is a fundamental solution to PDE above, then

$$u(y) = (\Lambda_F * f)(y) = \langle \Lambda_F, \tau_y \tilde{f} \rangle = \int F(x)f(y-x)dx = \int F(y-x)f(x)dx$$

So we have $u(y) = \int G(x,y)f(x)dx$ where $G(x,y) = F(y-x)$.

What about initial/ boundary conditions? What about homogeneous equations?

$$u(y) = \beta(y) + \int_K G(x,y)f(x)dx$$

Risks and Kernels

Empirical Risk and Empirical Regularized Risk:

$$\hat{R}(\beta, G) = \frac{1}{N} \sum_{i=1}^{N} \left|\left| u_i - \beta - \int_{D_f} G(x, \cdot) f_i(x) dx \right|\right|^2_{L^2(D_u)}$$

$$\hat{R}_{\rho,\lambda}(\beta, G) = \hat{R}(\beta, G) + \rho ||\beta||_{\mathcal{B}} + \lambda ||G||_{\mathcal{G}}$$

Goal:

$$\hat{\beta}_{N,\rho,\lambda}, \hat{G}_{N,\rho,\lambda} =_{\beta \in \mathcal{B}, G \in \mathcal{G}} \hat{\mathcal{R}}_{\rho,\lambda}(\beta, G)$$

Exponential Kernels:

$$K(x, y, \xi, \eta) = \exp\left(-\sqrt{\frac{||x - \xi||^2}{\sigma_x^2} + \frac{||y - \eta||^2}{\sigma_y^2}} \right)$$

$$Q(y, \eta) = \exp\left(-\frac{||y - \eta||}{\sigma_y} \right)$$

Representer Theorem for Green's Functions

Theorem

For optimal $\hat{\beta}_{N,\rho,\lambda}$ and $\hat{G}_{N,\rho,\lambda}$ on data $\{f_i, u_i\}_{i=1}^n \subset L^2(D_f) \times L^2(D_u)$, the function $\hat{G}_{N,\rho,\lambda}$ must have the form

$$\hat{G}_{N,\rho,\lambda}(x,y) = \sum_{i=1}^{n} \int_{D_f \times D_u} K(x,y,\xi,\eta) f_i(\xi) c_i(\eta) d(\xi,\eta)$$

where $c_i \in L^2(D_u)$ for all $i \in \{1, 2, \ldots, n\}$

- Model $\mathcal{G}$ as a space of linear operators $\mathcal{O}$ with $O_G(f) = y \mapsto \int_{D_f} G(x,y) f(x) dx$ and define $\langle O_G, O_{G'} \rangle = \langle G, G' \rangle_{\mathcal{G}}$
- Define an operator-valued kernel $\mathcal{K} : L^2(D_f) \times L^2(D_f) \to \mathcal{L}(L^2(D_u))$

$$[\mathcal{K}(f,g)u](y) = \int_{D_f} \int_{D_f} \int_{D_u} K(x,y,\xi,\eta) g(x) f(\xi) u(\eta) d\eta d\xi dx$$

for $f, g \in L^2(D_f)$ and $u \in L^2(D_y)$

Green's Function Approach to Solving Linear PDEs

Target PDE (e.g. Schrödinger Equation):

$$\left(-\frac{1}{2}\nabla^2 + V(x)\right) u(x) = f(x)$$

Green's Function Representation:

$$u(x) = \int G(x,y) f(y)\, dy \quad \text{where } \mathcal{L}_x G(x,y) = \delta(x-y)$$

Stepaniants et al. (JMLR 2023):

- Represent $G(x,y) \in \mathcal{H}_K$, a Reproducing Kernel Hilbert Space
- Learn G from data $\{f_i, u_i\}$ using functional regression
- Loss function (empirical risk):

$$\hat{R}(G) = \frac{1}{N} \sum_{i=1}^{N} \left\| u_i - \int G(x,y) f_i(y)\, dy \right\|_{L^2}^2$$

- Add regularization:

$$\hat{R}_{\rho,\lambda}(G) = \hat{R}(G) + \lambda \|G\|_{\mathcal{H}_K}^2$$

Key Benefit: Green's function learning approximates $\mathcal{L}^{-1}$ directly from data.

From Linear to non-Linear PDEs

Linear vs Nonlinear PDEs:

Linear: $\mathcal{L}u = f \Rightarrow u = Gf$

Nonlinear: $\mathcal{L}[u]u = f \Rightarrow u$ depends on itself

Kohn–Sham DFT Equation (Nonlinear):

$$\left(-\tfrac{1}{2}\nabla^2 + V_{\text{eff}}[\rho]\right) u = \varepsilon u, \quad \rho = |u|^2$$

$$V_{\text{eff}}[\rho] = V_{\text{ext}} + V_H[\rho] + V_{\text{XC}}[\rho]$$

1D H_2-like System (soft Coulomb):

$$V_{\text{ext}}(x) = -\frac{1}{\sqrt{(x-R_1)^2+\epsilon}} - \frac{1}{\sqrt{(x-R_2)^2+\epsilon}}$$

$$V_H(x) = \int \frac{\rho(x')}{|x-x'|+\epsilon} dx'$$

$$V_{\text{XC}}(x) = -0.75 \cdot \rho(x)^{1/3}$$

Takeaway: Nonlinearity prevents closed-form Green's function $\rightarrow$ iterative solution (e.g., SCF) required.

Motivation and Solver Comparison

Goal: Can RKHS-learned Green's function $G(x, y)$ replace standard DFT solvers?

Experimental Steps:

1. Set up a 1D H2-like Kohn–Sham system
2. Generate reference solution using traditional SCF method
3. Compare three solver types

Solver Comparison

Method	**workflow**	**Characteristics**
SCF (Standard)	Iterative: $\rho \to V_{\text{eff}}[\rho] \to u \to \rho$	+ Accurate – Slow convergence
Spectral	Basis expansion: $u(x) = \sum c_n \phi_n(x)$, solve algebraically	+ Guaranteed convergence, interpretable – Depends on basis quality
RKHS (Ours)	Learn $G(x, y)$ from (f_i, u_i), apply via $u = Gf$	+ Avoids solving PDEs at inference – unstable convergence, black-box

Evaluation:

$$u(x), \quad \rho(x) = |u(x)|^2, \quad \text{Total energy}, \quad \text{Convergence}$$

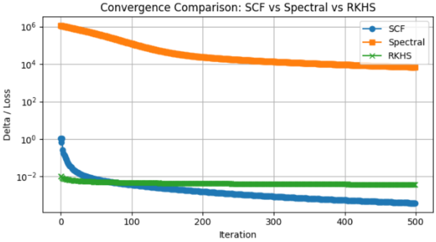
Results

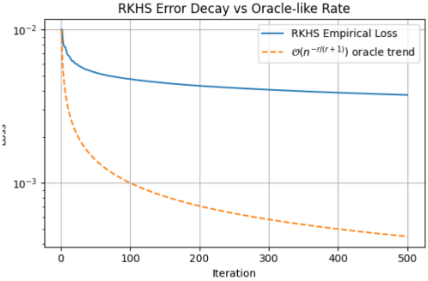
Key Observations

- RKHS method can learn a structured Green's function $G(x, y)$ from data in each step.
- Convergence speed: SCF(tradition) $>$ RKHS $>$ Spectral

Oracle Inequality-inspired Decay

$$R(\hat{G}_n) - R(G^*) = \mathcal{O}\left(\frac{1}{n^{r/(r+1)}}\right), \quad \text{(theoretical)}$$





Critical Analysis and Outlook

- Limitations:
  - Only linear PDEs currently addressed
  - Kernel choice impacts stability and expressiveness
  - Computational cost scales with grid size and number of energies (In some cases, the original method might be faster)
- Extensions:
  - Apply to higher-dimensional problems
  - Learn nonlinear operators via operator-valued kernels
  - Integrate into full DFT package with SCF iterations

Appendix: Point Source- Dirac Delta Function

Definition

The Dirac delta function $\delta(x)$ is a "distribution" such that for all smooth functions f,

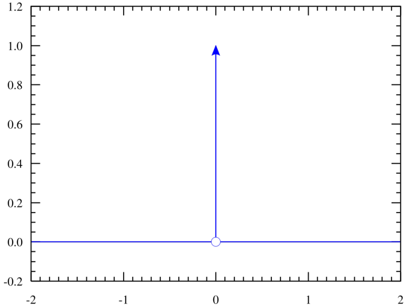
$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)\, dx = f(x_0).$$

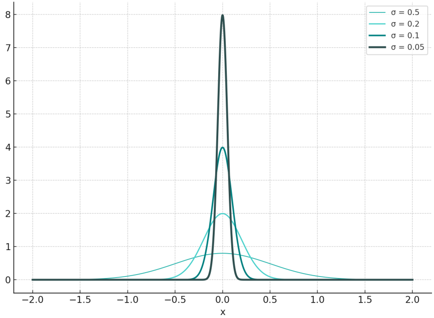
Heuristically, we can say

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases}, \quad \int_{-\infty}^{\infty} \delta(x) dx = 1.$$

- In a sloppy sense, Dirac Delta can serve as a positive semi-definite kernel of RKHS, $k(x, x') \approx \delta(x - x')$.
- Many representations exist, e.g., $\delta(x - x') = \lim_{\sigma \to 0} \mathcal{N}(x', \sigma)$.

Appendix: Point Source- Dirac Delta Function





$$\delta(x)$$

$$\delta_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Appendix: $-\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|} = \delta^3(\vec{x}-\vec{x}')$

Let $\vec{r} = \vec{x} - \vec{x}'$ and $r = |\vec{r}|$. $\vec{\nabla} = \vec{\nabla}_x = \sum_i \hat{x}_i \partial_{x_i}$.
Note that

$$\vec{\nabla}^2 \frac{1}{|\vec{x}-\vec{x}'|} = \vec{\nabla}^2 \frac{1}{r} = \vec{\nabla} \cdot \left(\frac{-\hat{r}}{r^2} \right) = \vec{\nabla} \cdot \left(\frac{-\vec{r}}{r^3} \right). \tag{4}$$

and

$$\vec{\nabla} \cdot (f(r)\ \vec{r}) = \partial_x (f(r)x) + \partial_y (f(r)y) + \partial_z (f(r)z) \tag{5}$$

$$= 3f(r) + (\partial_x r + \partial_y r + \partial_z r)(df/dr) = 3f(r) + r\frac{df}{dr}. \tag{6}$$

Then, for $f(r) = r^{n-1}$, $\vec{\nabla} \cdot (r^{n-1}\vec{r}) = \vec{\nabla} \cdot (r^n \hat{r}) = (n+2)r^{n-1}$, which vanishes for $n = -2$, except at $r = 0$.

Appendix: $-\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|} = \delta^3(\vec{x}-\vec{x}')$

For $\vec{x}-\vec{x}' \neq \vec{0}$, $-\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|} = \vec{\nabla} \cdot \left(\frac{\vec{x}-\vec{x}'}{4\pi|\vec{x}-\vec{x}'|^3} \right) = 0$ (previous page).

For $\vec{x}-\vec{x}' \to \vec{0}$, consider a small sphere with radius ε centered at $\vec{x}'$.

Do volume integration of $-\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|}$ in the sphere:

$$\int_V -\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|} d^3x = \int_V \vec{\nabla} \cdot \left(\frac{\vec{x}-\vec{x}'}{4\pi|\vec{x}-\vec{x}'|^3} \right) d^3x \tag{7}$$

$$= \int_S \frac{\vec{x}-\vec{x}'}{4\pi|\vec{x}-\vec{x}'|^3} \cdot d\vec{\sigma} \quad \text{Divergence theorem} \tag{8}$$

$$= \int_S \frac{\hat{r}}{4\pi\varepsilon^2} \cdot \hat{r} dS = \frac{1}{4\pi\varepsilon^2} \int_S dS \tag{9}$$

$$= \frac{1}{4\pi\varepsilon^2} 4\pi\varepsilon^2 = 1 \text{ for arbitrary small } \varepsilon. \tag{10}$$

So, $-\vec{\nabla}^2 \frac{1}{4\pi|\vec{x}-\vec{x}'|}$ is a Dirac Delta function in 3-d Euclidean space.