Generalized Kohn–Sham iteration on Banach spaces

Andre Laestadius*

Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry,
University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

Markus Penz

Max Planck Institute for the Structure and Dynamics of Matter,
Luruper Chausse 149, 22761 Hamburg, Germany

Erik I. Tellgren

Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry,
University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

Michael Ruggenthaler

Max Planck Institute for the Structure and Dynamics of Matter,
Luruper Chausse 149, 22761 Hamburg, Germany

Simen Kvaal

Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry,
University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway

Trygve Helgaker

Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry,
University of Oslo, P.O. Box 1033 Blindern, N-0315 Oslo, Norway and
Centre for Advanced Study at the Norwegian Academy of Science and Letters, Drammensveien 78, N-0271 Oslo, Norway

(Dated: May 1, 2018)

A detailed account of the Kohn–Sham algorithm from quantum chemistry, formulated rigorously in the very general setting of convex analysis on Banach spaces, is given here. Starting from a Levy–Lieb-type functional, its convex and lower semi-continuous extension is regularized to obtain differentiability. This extra layer allows to rigorously introduce, in contrast to the common unregularized approach, a well-defined Kohn–Sham iteration scheme. Convergence in a weak sense is then proven. This generalized formulation is applicable to a wide range of different density-functional theories and possibly even to models outside of quantum mechanics.

I. INTRODUCTION

Density-functional theory (DFT) is usually presented as based on the Hohenberg–Kohn theorem and received a thorough mathematical investigation in terms of convex analysis by Lieb [1]. Yet it is the Kohn–Sham (KS) iteration scheme [2], built upon the consequences of the Hohenberg–Kohn theorem, that makes it applicable to problems of quantum chemistry and where it developed an unprecedented utility. In Kohn–Sham theory a complex, interacting system with basic state variable x (density) is compared against a simpler, non-interacting reference system that reproduces the exact same state x. To achieve that, the reference system is complemented with an auxiliary potential x^*_{KS} that acts as a dual system variable and gets fixed by the connection of both systems via their respective energy functionals. Since no explicit expression for the Kohn–Sham potential x^*_{KS} is at hand, a clever self-consistent iteration scheme is set up (see Section III B).

Nonetheless, up until now very few results about the convergence of this procedure are at hand. Among mathematical analyses of the self-consistent field procedure used to solve the Hartree–Fock or Kohn–Sham equations, we want to highlight in particular the work on the optimal damping algorithm (ODA) formulated in terms of one-particle reduced density matrices and Kohn–Sham matrices [3–5].

In a setting where computationally efficient, approximate density functionals are used, the Kohn–Sham wave function as the underlying physical structure is crucial for providing a good first approximation to the kinetic energy, since this is, in practice, very hard to approximate from the density alone. However, in the rigorous mathematical setting of this work, where properties of the exact density functionals are analyzed and employed, the Kohn–Sham wave function itself does not show up. Somewhat unconventionally, one can therefore choose to work with the exact energy of the Kohn–Sham system as a (orbital-free) density functional. This makes the whole analysis just as much applicable to orbital-free methods like Thomas–Fermi theory. The analogue of the self-consistent field procedure in this setting is then

* andre.laestadius@kjemi.uio.no

a procedure that determines a sequence of densities and Kohn–Sham potentials that converges to a self-consistent pair. Working in this setting, Wagner et al. [6] adapted the ODA and claimed, unfortunately mistakenly, to show convergence to a self-consistent pair of a ground-state density and a Kohn–Sham potential. Furthermore, the usual setting of Kohn–Sham theory is largely built on ill-defined quantities (cf. Section II A). Only more recent work by Kvaal et al. [7] showed for the first time how the so-called Moreau–Yosida regularization of convex functionals, explained in Section II F, can be employed to set up the iteration scheme in a well-defined manner on Hilbert spaces for density and potential variables. However, the important problem of convergence was not addressed in this mathematically strict formulation of the theory.

In this work we aim at closing gaps of Kohn–Sham theory while at the same time considerably extending its scope. We both (i) generalize the regularized Kohn–Sham iteration scheme of Kvaal et al. [7] to Banach spaces and (ii) show a weak form of convergence. Flaws of the previous convergence proof in Wagner et al. [6] that we repeat for the well-defined setting of regularized functionals are uncovered in Section III B, Remark 4. Such a generalized Banach space formulation of DFT has the advantage of laying possible solid foundations to many types of ground-state DFTs, including more recent developments in current-density-functional theory [8], quantum-electrodynamical DFT [9], thermal DFT [10], and reduced density matrix functional theory [11, 12].

Note that, even though most of the presentation has a link to quantum mechanics, the formulation here is kept in a general fashion that might prove valuable for applications to other fields. The usual DFT setting for many-electron quantum systems is presented in the final section Section IV where we also discuss extensions of the standard theory that include magnetic fields.

II. PRELIMINARIES

A. The general problem

The principal state variable x (density) is chosen from a real Banach space X. The dual space of X (consisting of potentials) is

$$X^* = \{x^* : X \to \mathbb{R} \mid x^* \text{ linear and continuous}\}$$

and we write $\langle x^*, x \rangle = x^*(x)$ for the dual pairing between $x \in X$ and $x^* \in X^*$. Additional conditions for these spaces will later include strict convexity for both X and X^* and uniform convexity for the dual space X^*. Note that uniform convexity of X implies both reflexivity of X and X^* (Milman–Pettis theorem) as well as strict convexity of X^*. By a theorem due to Asplund, if X is assumed reflexive, an equivalent norm can be chosen such that X and X^* are strictly convex (yet not uniformly convex). Important uniformly convex spaces are

the Lebesgue spaces L^p with $1 < p < \infty$ but not L^1 or L^∞. In a strictly convex space the connecting line segment between surface points of the unit ball lies strictly inside the ball, while in a uniformly convex space the distance of the middle point of the line segment from the surface is not only non-zero but also depenends only on the length of the segment, not on the chosen points themselves. (For further details see [13, Section 1.2.4].)

In DFT, one is studying a variational problem that consists of finding the N-electron system state corresponding to minimal energy (see Section IV A). For the general problem we consider an energy functional $E : X^* \to \mathbb{R}$ given by

$$E(x^*) = \inf\{\tilde{F}(x) + \langle x^*, x \rangle \mid x \in \tilde{X}\}, \tag{1}$$

for some functional $\tilde{F} : \tilde{X} \to \mathbb{R}$, $\tilde{X} \subseteq X$, that originates from the underlying physical problem. In such models $\tilde{F}(x)$ stands for all energy contributions of internal effects, while the dual pairing $\langle x^*, x \rangle$ represents the potential energy that is seen as an external and controllable effect. The domain of $\tilde{F}$ is limited to a certain set of "physical" densities $\tilde{X}$ that by themselves usually do not form a linear space and $\tilde{F}(x)$ represents the minimal possible internal energy for a given state x. Such a $\tilde{F}$, called the Levy–Lieb functional [1, 14] in the DFT literature, is in general not convex or lower semi-continuous. Therefore one introduces (borrowing terminology from DFT) the universal Lieb functional, $F : X \to \mathbb{R} \cup \{+\infty\}$, as [1]

$$F(x) = \sup\{E(x^*) - \langle x^*, x \rangle \mid x^* \in X^*\}, \tag{2}$$

in terms of which we may obtain the energy as

$$E(x^*) = \inf\{F(x) + \langle x^*, x \rangle \mid x \in X\}. \tag{3}$$

The functional F is by construction convex and lower semi-continuous [1, Theorem 3.6] and has $F(x) = +\infty$ whenever $x \in X \setminus \tilde{X}$ (corresponding to "unphysical" densities). If one chooses $x^* \in X^*$ such that a minimizer $x \in \tilde{X}$ for (1) exists and thus $E(x^*) = \tilde{F}(x) + \langle x^*, x \rangle$, then it follows after insertion into (2) that $F(x) = \tilde{F}(x)$. Each minimizer in (1) is therefore a minimizer in (3) but the converse does not hold. A minimizer in (3) need not be a minimizer of (1) and in general $F \leq \tilde{F}$ on $\tilde{X}$.

To find the state x of minimal energy in (1) by just relying on F, we must determine

$$\arg\min\{F(x) + \langle x^*, x \rangle \mid x \in X\}, \tag{4}$$

calling for differentiation of $F(x) + \langle x^*, x \rangle$ with respect to x. But even though F already has some nice properties, one cannot assume that F is differentiable [15]. Yet the more general notion of "subdifferential" (see Definition 2) gives a non-unique, implicit answer to (4), namely

$$-x^* \in \underline{\partial} F(x) \iff E(x^*) = F(x) + \langle x^*, x \rangle. \tag{5}$$

For a given density x, a non-empty subdifferential means that this state is "v-representable" (we will use this term

commonly found in the literature instead of calling it "x^*-representable"), meaning that it can be found as the minimizer to (1) with a corresponding potential in X^*. (Note that the concept of v-representability depends on the choice of X and X^*.) For given x^*, (5) is the variational problem that finally gets approximated with the aid of a Kohn–Sham reference system. To facilitate the iteration scheme with well-defined quantities, Moreau–Yosida regularization (see Section II F) is employed. This yields a functional F_ε that is also functionally differentiable. But before that, we give the relation between the fundamental functionals E and F a strict mathematical meaning.

B. Convex conjugates

To start with, we say that a convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ is proper if not identical to $+\infty$. Let $\Gamma_0(X)$ denote the set of proper, convex, lower semi-continuous functions $X \to \mathbb{R} \cup \{+\infty\}$. On the dual side, $\Gamma_0^*(X^*)$ is the set of proper convex and weak-* lower semi-continuous functions $X^* \to \mathbb{R} \cup \{+\infty\}$. We also introduce the sets $\Gamma(X) = \Gamma_0(X) \cup \{\pm\infty\}$ and $\Gamma^*(X^*) = \Gamma_0^*(X^*) \cup \{\pm\infty\}$. Note that $\Gamma(X)$ does not contain all lower semi-continuous convex functions on X, only all proper lower semi-continuous convex functions and the improper convex functions $\pm\infty$. Collectively, these functions are known as the closed convex functions on X. Likewise, $\Gamma_0^*(X^*)$ contains the closed convex functions on X^*.

Key in the presented framework will be two close relatives to the convex conjugate (Legendre–Fenchel transformation), a non-standard definition introduced in [1, 7]. For $f : X \to \mathbb{R} \cup \{\pm\infty\}$ and $g : X^* \to \mathbb{R} \cup \{\pm\infty\}$, we define the (skew) conjugate functions:

$$f^\wedge(x^*) = \inf\{f(x) + \langle x^*, x\rangle \mid x \in X\} \in -\Gamma^*(X^*),$$
$$g^\vee(x) = \sup\{g(x^*) - \langle x^*, x\rangle \mid x^* \in X^*\} \in \Gamma(X).$$

The following non-trivial result relates f and g to their biconjugates $f = (f^\wedge)^\vee$ and $g = (g^\vee)^\wedge$, respectively:

Theorem 1. *The Fenchel–Moreau biconjugation theorem states that*

$$f = (f^\wedge)^\vee \iff f \in \Gamma(X),$$
$$g = (g^\vee)^\wedge \iff g \in -\Gamma^*(X^*).$$

This theorem sets up a one-to-one correspondence between the closed convex functions on X and the closed concave functions on X^*. If f and g are not closed convex and concave, respectively, then the weaker results $f \geq (f^\wedge)^\vee$ and $g \leq (g^\vee)^\wedge$ hold and we may think of biconjugation as performing "gamma regularizations" of f and g.

We have already used the skew conjugations in (1)–(3) above, which may now be more succinctly written in the

form (having extended $\tilde{F}$ to to the whole X by setting it equal to $+\infty$ outside $\tilde{X}$):

$$E = \tilde{F}^\wedge, \quad F = E^\vee, \quad E = F^\wedge.$$

We note that the ground-state energy is closed concave $E \in -\Gamma_0^*(X^*)$ and that the closed convex Lieb functional $F = (\tilde{F}^\wedge)^\vee \in \Gamma_0(X)$ is the gamma regularization of the Levy–Lieb functional $\tilde{F} \notin \Gamma(X)$. (In the DFT context, the Lieb functional is the gamma regularization of any admissible density functional in the Hohenberg–Kohn variational principle.)

C. Banach space derivatives

Definition 2. We give the following notions for derivatives of a function $f : X \to \mathbb{R} \cup \{+\infty\}$:

(i) f is Gâteaux differentiable at $x \in \mathrm{dom}(f)$ if there exists a $\nabla f(x) \in X^*$ such that for all $u \in X$

$$\lim_{t\to 0} \frac{1}{t}(f(x + tu) - f(x)) = \langle \nabla f(x), u\rangle.$$

(ii) f is Fréchet differentiable at $x \in \mathrm{dom}(f)$ if there exists a $\nabla f(x) \in X^*$ such that

$$\lim_{h\to 0} \|h\|^{-1}(f(x + h) - f(x) - \langle \nabla f(x), h\rangle) = 0.$$

(iii) f is subdifferentiable at $x \in X$ if it has a nonempty subdifferential $\underline{\partial}f(x) \subset X^*$ at x given by

$$\underline{\partial}f(x) = \{x^* \in X^* \mid \forall y \in X :$$
$$f(x) - f(y) \leq \langle x^*, x - y\rangle\}.$$

The elements of $\underline{\partial}f(x)$ are known as the subgradients of f at x.

We denote by $\mathrm{dom}(f)$ the effective domain of f (the subset of X where f is finite) and by $\mathrm{dom}(\underline{\partial}f)$ the domain of subdifferentiability of f (the subset of X where $\underline{\partial}f$ is non-empty). We note that $\mathrm{dom}(\underline{\partial}f) \subset \mathrm{dom}(f)$ but $\underline{\partial}f(x) = \emptyset$ may happen also for $x \in \mathrm{dom}(f)$. If f is closed convex, then $\mathrm{dom}(\underline{\partial}f)$ is dense in $\mathrm{dom}(f)$ by the Brøndsted–Rockafellar theorem.

Subdifferentials are defined for arbitrary functions but are particularly useful for convex functions. On the dual space, we mostly work with concave functions, whose superdifferentials are defined by

$$\overline{\partial}g(x^*) = \{x \in X \mid \forall y \in X^* :$$
$$g(x^*) - g(y^*) \geq \langle x^* - y^*, x\rangle\}.$$

Note that $\overline{\partial}g(x^*)$ is by definition a subset of X rather than X^{**}. In an obvious manner, we may also define supergradients on X and subgradients on X^*.

D. Optimality conditions

Sub- and superdifferentiation are precisely the tools needed to characterize optimality of convex and concave conjugations. The following results are easily established from the definition of the sub- and superdifferentials:

Lemma 3. *If $f : X \to \mathbb{R} \cup \{\pm\infty\}$ and $g : X^* \to \mathbb{R} \cup \{\pm\infty\}$, then*

$$-x^* \in \underline{\partial}f(x) \iff f^\wedge(x^*) = f(x) + \langle x^*, x\rangle,$$
$$x \in \overline{\partial}g(x^*) \iff g^\vee(x) = g(x^*) + \langle x^*, x\rangle.$$

Proof. From the definitions of subdifferential and conjugate function, we have

$$-x^* \in \underline{\partial}f(x)$$
$$\iff \forall y \in X^* : f(y) + \langle x^*, y\rangle \geq f(x) + \langle x^*, x\rangle$$
$$\iff \inf\{f(y) + \langle x^*, y\rangle \mid y \in X^*\} \geq f(x) + \langle x^*, x\rangle$$
$$\iff \inf\{f(y) + \langle x^*, y\rangle \mid y \in X^*\} = f(x) + \langle x^*, x\rangle$$
$$\iff f^\wedge(x^*) = f(x) + \langle x^*, x\rangle.$$

The result for g follows in a similar manner. $\square$

For closed convex functions, we obtain by setting $g = f^\wedge$ in Lemma 3 and applying the biconjugation theorem:

Lemma 4. *[13, Proposition 2.33] For $f \in \Gamma_0(X)$ we have*

$$-x^* \in \underline{\partial}f(x) \iff x \in \overline{\partial}f^\wedge(x^*).$$

Finding a solution to $-x^* \in \underline{\partial}f(x)$ is therefore equivalent to finding a solution to $x \in \overline{\partial}f^\wedge(x^*)$. This switch between primal and dual problems will be used several times in throughout this work.

E. Monotonicity of subdifferentials

Lemma 5. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be proper convex. Then $\underline{\partial}f$ is a monotone operator, that is, for any $x, y \in X$, each pair $x^* \in \underline{\partial}f(x)$ and $y^* \in \underline{\partial}f(y)$ satisfies*

$$\langle x^* - y^*, x - y\rangle \geq 0.$$

If f is strictly convex, then $\underline{\partial}f$ is strictly monotone, meaning that the inequality holds strictly for $x \neq y$. For concave functions, the inequality is reversed.

Proof. Let $x, y \in X$ be subdifferentiable points and select subgradients $x^* \in \underline{\partial}f(x), y^* \in \underline{\partial}f(y)$. Then

$$f(x) - f(y) \leq \langle x^*, x - y\rangle,$$
$$f(y) - f(x) \leq \langle y^*, y - x\rangle = \langle -y^*, x - y\rangle.$$

Adding these subgradient inequalities together, we obtain monotonicity. If f is strictly convex, then the subgradient inequalities hold strictly for $x \neq y$, yielding strict monotonicity. $\square$

The above lemma is a rigorous convex analytical version of a more physically motivated lemma in Wagner et al. [6] and is equivalent to the Hohenberg–Kohn theorem. To see that, assume E strictly concave (which is not true generally but will be the case for the regularized version, see Remark 3) and take $x \in \overline{\partial}E(x^*)$, $y \in \overline{\partial}E(y^*)$. Then by Lemma 5 we get strict monotonicity of $\overline{\partial}E$, and with the roles of x, y and x^*, y^* interchanged, it follows from $x^* \neq y^*$ that $x \neq y$. This means that different potentials always lead to different states and that the corresponding mapping $x^* \mapsto x \in \overline{\partial}E(x^*)$ is injective and thus invertible on the restricted codomain of v-representable states. This is just the main statement of the Hohenberg–Kohn theorem (including possible degeneracies), which is discussed again in the standard setting of DFT in Section IV A.

F. Moreau–Yosida regularization

Definition 6. Set $\phi(\cdot) = \frac{1}{2}\|\cdot\|^2$ on X and X^*. Then for $\varepsilon > 0$ the Moreau–Yosida regularization of a $f \in \Gamma_0(X)$ is defined as

$$f_\varepsilon(x) = \inf\{f(y) + \varepsilon^{-1}\phi(x - y) \mid y \in X\}.$$

Remark 1. The function ϕ is strictly convex if and only if the respective Banach space is strictly convex [13, Proposition 1.103].

Remark 2. Alternatively we can define $f_\varepsilon = f \,\square\, \varepsilon^{-1}\phi$, where the box notation stands for the *infimal convolution* of f and g and is given by

$$(f \,\square\, g)(x) = \inf\{f(y) + g(x - y) \mid y \in X\}.$$

A minimal value of f (if it exists) is preserved at the same location when passing over to f_ε. The infimum in the definition above is always uniquely attained, $f_\varepsilon \in \Gamma_0(X)$ and is everywhere finite on X [13, §2.2.3]. This unique minimizer for any given x gives rise to the definition of the proximal mapping

$$\operatorname*{prox}_{\varepsilon f}(x) = \arg\min\{f(y) + \varepsilon^{-1}\phi(x - y) \mid y \in X\}$$
$$= \arg\min\{\varepsilon f(y) + \phi(x - y) \mid y \in X\}.$$

Definition 7. The duality map on X is

$$J(x) = \{x^* \in X^* \mid \|x^*\| = \|x\|, \langle x^*, x\rangle = \|x\|^2\} \quad (6)$$

and assigns to each state a set of dual elements.

The set $J(x)$ is a singleton if and only if X^* is strictly convex, J is injective if and only if X is strictly convex, and J is surjective if and only if X is reflexive [13, Proposition 1.117]. Of particular interest is the subdifferential of ϕ that yields just the duality map [13, Example 2.32]

$$\underline{\partial}\phi(x) = J(x). \quad (7)$$

Under the assumption that X is reflexive, we can define J^{-1} on all of X^* and then have (7) also in the reverse direction

$$\underline{\partial}\phi(x^*) = J^{-1}(x^*). \tag{8}$$

Proposition 8. *[16] If X^* is strictly convex then ϕ is Gâteaux differentiable, if X^* is uniformly convex then ϕ is Fréchet differentiable. In both cases the derivative is the duality map J.*

If we recall that the Moreau–Yosida regularization of a functional corresponds to the infimal convolution with ϕ (Remark 2), we can expect that the regularity properties of ϕ that follow from Proposition 8 are taken over to the regularized functional $f_\varepsilon = f \Box \varepsilon^{-1}\phi$. This means f_ε should be Gâteaux differentiable if X^* is strictly convex and Fréchet differentiable if X^* is uniformly convex, statements that get proven by the following theorem.

Theorem 9. *Suppose $f \in \Gamma_0(X)$ and that X, X^* are strictly convex. Then f_ε is Gâteaux differentiable on X and*

$$\nabla f_\varepsilon(x) = \varepsilon^{-1} J(x - \operatorname*{prox}_{\varepsilon f}(x)).$$

If X^* is uniformly convex, then f_ε is even Fréchet differentiable.

Proof. The proof of the first part can be found in [13], Theorem 2.58, with the derivative given by (2.57) there. That the derivative can indeed be evaluated by application of the proximal mapping is shown in the proof of Theorem 2.58. Proposition 1.146 (i) in [13] further states that if X^* is uniformly convex then $\nabla f_\varepsilon : X \to X^*$ is continuous, which implies Fréchet differentiability [17, Lemma 34.3]. $\square$

We now turn our attention to the convex conjugate pair $E = F^\wedge$ from (1) and (2) and their regularized versions

$$F_\varepsilon(x) = \inf\{F(y) + \varepsilon^{-1}\phi(x - y) \mid y \in X\},$$
$$E_\varepsilon(x^*) = (F_\varepsilon)^\wedge(x^*).$$

Note especially that E_ε is not the regularization of E but rather the transformation of the regularized F_ε.

Theorem 10. *Suppose $F \in \Gamma_0(X)$ and X reflexive then*

$$E_\varepsilon(x^*) = E(x^*) - \varepsilon\phi(x^*) \quad and \tag{9}$$
$$\overline{\partial}E_\varepsilon(x^*) = \overline{\partial}E(x^*) - \varepsilon J^{-1}(x^*). \tag{10}$$

Proof. To prove (9), note that by definition

$$
\begin{aligned}
E_\varepsilon(x^*) &= (F_\varepsilon)^\wedge(x^*)\\
&= \inf\{F(y) + \varepsilon^{-1}\phi(x - y) + \langle x^*, x\rangle \mid x, y \in X\}\\
&= \inf\{F(y) + \langle x^*, y\rangle\\
&\quad + \varepsilon^{-1}\phi(x - y) + \langle x^*, x - y\rangle \mid x, y \in X\}\\
&= \inf\{F(y) + \langle x^*, y\rangle \mid y \in X\}\\
&\quad + \inf\{\varepsilon^{-1}\phi(z) + \langle x^*, z\rangle \mid z \in X\}\\
&= F^\wedge(x^*) + (\varepsilon^{-1}\phi)^\wedge(x^*).
\end{aligned}
$$

From $\phi(x) = \|x\|^2/2$ it follows that $\phi^\wedge(x^*) = -\|x^*\|^2/2$. Furthermore, the scaling relation $(\lambda f)^\wedge(x^*) = \lambda f^\wedge(x^*/\lambda)$, $\lambda > 0$, gives $(\varepsilon^{-1}\phi)^\wedge(x^*) = -\varepsilon\|x^*\|^2/2$. Thus

$$(\varepsilon^{-1}\phi)^\wedge(x^*) = -\varepsilon\phi(x^*)$$

and we can conclude that (9) holds. Eq. (10) follows directly from forming the superdifferential of (9) and inserting (8). $\square$

Remark 3. Eq. (9) shows that E_ε is *strictly* concave as the sum of a concave function and the strictly concave $-\varepsilon\phi$. From strict concavity follows that $E_\varepsilon(x^*) - \langle x^*, x\rangle$ attains its maximum at one point only, so the regularized version of (2) has a unique maximizer. The subdifferential of the conjugate $F_\varepsilon = (E_\varepsilon)^\vee$, which gives this maximizer just like in (5), is thus a singleton. That the subdifferential ∂F_ε gives a singleton is naturally true in the case of Gâteaux differentiability as in Theorem 9. (See [13] just above Proposition 2.33 and also note Proposition 2.40 in this context.) Strict concavity of the energy functional also connects to strict monotonicity of its superdifferential and the Hohenberg–Kohn theorem, see Section II E.

Corollary 11. *Let X and X^* be strictly convex and reflexive. Any solution of the regularized problem $x \in \overline{\partial}E_\varepsilon(x^*)$ is connected to a "physical" solution of the unregularized problem by the proximal mapping*

$$\operatorname*{prox}_{\varepsilon F}(x) \in \overline{\partial}E(x^*) \subset \tilde{X}.$$

Proof. Take $x \in \overline{\partial}E_\varepsilon(x^*)$, which is equivalent to $-x^* \in \underline{\partial}F_\varepsilon(x)$ due to Lemma 4. This subdifferential is even a singleton by Theorem 9 and it holds

$$-x^* = \varepsilon^{-1} J(x - \operatorname*{prox}_{\varepsilon F}(x)).$$

Since the Banach space is strictly convex and reflexive, the duality map is bijective and we apply J^{-1} to get

$$-\varepsilon J^{-1}(x^*) = x - \operatorname*{prox}_{\varepsilon F}(x).$$

We have here also used the fact that the duality map is always a homogeneous function. Now, comparing this with (10) in Theorem 10, we conclude from $x \in \overline{\partial}E_\varepsilon(x^*)$ that $\operatorname{prox}_{\varepsilon F}(x) \in \overline{\partial}E(x^*)$. It finally follows from the definition of E in (1) that this solution is always in $\tilde{X}$. $\square$

III. REGULARIZED KOHN–SHAM ITERATION

A. Basic setup

Assume that $x^* \in X^*$ has been given and we want to obtain the ground-state energy $E(x^*)$ from (1) and

a corresponding minimizer x from (4), meaning that we must satisfy the equivalent conditions

$$-x^* \in \underline{\partial}F(x) \iff x \in \overline{\partial}E(x^*).$$

Transforming to the regularized energy functional with the help of Theorem 10, we obtain

$$x \in \overline{\partial}E(x^*) = \overline{\partial}E_\varepsilon(x^*) + \varepsilon J^{-1}(x^*),$$
$$E(x^*) = E_\varepsilon(x^*) + \varepsilon\phi(x^*).$$

Parallel to that we assume the existence of a reference functional $\tilde{F}^0 : \tilde{X} \to \mathbb{R}$ that belongs to the Kohn–Sham system and captures parts of the systems internal physics and leads to a variational problem that is supposedly easier to solve. Just like with $\tilde{F}$ we derive the regularized functionals F_ε^0 and E_ε^0 and set up the reference problem in analogous fashion

$$-x^*_{\text{KS}} \in \underline{\partial}F_\varepsilon^0(x) \iff x \in \overline{\partial}E_\varepsilon^0(x^*_{\text{KS}}),$$
$$-x^* \in \underline{\partial}F_\varepsilon(x) \iff x \in \overline{\partial}E_\varepsilon(x^*).$$

Note that the minimizer state x is the same in both cases, which makes it necessary to choose a different and at this stage undetermined potential for the reference system, the Kohn–Sham potential x^*_{KS}. The v-representability problem that strikes the standard Kohn–Sham construction at this point (not being able to choose a potential such that the *same* state is the solution) plays no role in the regularized version. By differentiability of F_ε and F_ε^0, achieved through regularization, such a potential always exists as the (negative) derivative of the regularized functionals

$$x^*_{\text{KS}} = -\nabla F_\varepsilon^0(x), \quad x^* = -\nabla F_\varepsilon(x).$$

Subtraction of those two equations yields the first step in a self-consistent iteration scheme that begins with a density x_1 as a guess to the minimizer x and eventually converges to x^*_{KS}. One sensible initial guess used in Theorem 12 later is $x_1 \in \overline{\partial}E_\varepsilon^0(x^*)$, which means just taking the fixed external potential as a rough first approximation to the Kohn–Sham potential also capturing internal effects. The next iteration towards x^*_{KS} is then given by

$$x^*_{i+1} = x^* + \nabla F_\varepsilon(x_i) - \nabla F_\varepsilon^0(x_i). \tag{11}$$

The expression $\nabla F_\varepsilon - \nabla F_\varepsilon^0$, although in practice not known explicitly, is at least accessible through approximations since major contributions are expected to cancel due to similar physical effects in both systems. In DFT, this potential is known by the name "Hartree exchange–correlation" and subsumes all interaction effects that are lost in the non-interacting reference system. The second part of the Kohn–Sham iteration is then the solution of the (simple or simpler) reference system

$$x_{i+1} \in \overline{\partial}E_\varepsilon^0(x^*_{i+1}).$$

The stopping condition is $x^*_{i+1} = -\nabla F_\varepsilon^0(x_i)$ because this gives $x^* = -\nabla F_\varepsilon(x_i)$ from (11), which means x_i is

the state yielding minimal (regularized) energy and consequently $x^*_{i+1} = x^*_{\text{KS}}$. Note that this Kohn–Sham potential belongs to the regularized reference system and the resulting state is "unphysical" and in general $x \notin \tilde{X}$. Effectively we introduced two transformational layers to the problem, firstly the reference system that gets connected to the sought-after solution with the whole Kohn–Sham procedure, and secondly the regularization. Invoking Theorem 10 and Corollary 11, it is possible to translate both the ground-state energy and the corresponding state back to the original unregularized layer.

The important question of Kohn–Sham convergence remains unaddressed up to this point. We shall see that, to guarantee at least weak convergence, the iteration must be slightly changed and not the full step from x_i to x_{i+1} is to be taken (following the ODA of the extended KS scheme in [5]).

B. Kohn–Sham iteration scheme

We can now formulate the iteration scheme and prove a weak form of convergence in terms of the energy.

Theorem 12. *Let X be strictly convex, X^* uniformly convex, $x^* \in X^*$ fixed, and set $x_1^* = x^*$ and $x_1 \in \overline{\partial}E_\varepsilon^0(x^*)$. Iterate $i = 1, 2, \ldots$ according to:*

(a) *Set $x^*_{i+1} = x^* + \nabla F_\varepsilon(x_i) - \nabla F_\varepsilon^0(x_i)$ and stop if $x^*_{i+1} = -\nabla F_\varepsilon^0(x_i) = x^*_{\text{KS}}$.*

(b) *Select $x'_{i+1} \in \overline{\partial}E_\varepsilon^0(x^*_{i+1})$.*

(c) *Choose $t_i \in (0, 1]$ maximally such that for $x_{i+1} = x_i + t_i(x'_{i+1} - x_i)$ one has*

$$\langle \nabla F_\varepsilon(x_{i+1}) + x^*, x'_{i+1} - x_i \rangle \leq 0.$$

Then the strictly descending sequence $\{F_\varepsilon(x_i) + \langle x^*, x_i \rangle\}_i$ converges as a sequence of real numbers to

$$e_\varepsilon(x^*) = \inf_i \{F_\varepsilon(x_i) + \langle x^*, x_i \rangle\} \geq E_\varepsilon(x^*).$$

Thus, the shifted $e_\varepsilon(x^*) + \varepsilon\phi(x^*)$ is an upper bound for the ground-state energy $E(x^*)$.

Proof. We note that F_ε and F_ε^0 are both Fréchet differentiable by Theorem 9 and start by checking the directional derivative of $F_\varepsilon + x^*$ at x_i in the step direction $x'_{i+1} - x_i$. From (a) in the KS scheme above, we have

$$\langle \nabla F_\varepsilon(x_i) + x^*, x'_{i+1} - x_i \rangle$$
$$= \langle x^*_{i+1} + \nabla F_\varepsilon^0(x_i), x'_{i+1} - x_i \rangle.$$

If $x^*_{i+1} = -\nabla F_\varepsilon^0(x_i)$ then from (a) $x^* = -\nabla F_\varepsilon(x_i)$, which is the desired ground-state solution, and we have converged to the KS potential $x^*_{\text{KS}} = x^*_{i+1}$. Otherwise, because of $x'_{i+1} \in \overline{\partial}E_\varepsilon^0(x^*_{i+1})$ and $x_i \in \overline{\partial}E_\varepsilon^0(-\nabla F_\varepsilon^0(x_i))$,

we can invoke Lemma 5 for the strictly concave E_ε^0 (see Remark 3) and get

$$\langle \nabla F_\varepsilon(x_i) + x^*, x'_{i+1} - x_i \rangle$$
$$= \langle x^*_{i+1} + \nabla F_\varepsilon^0(x_i), x'_{i+1} - x_i \rangle < 0.$$

Since F_ε is Fréchet differentiable, which means $\nabla F_\varepsilon : X \to X^*$ continuous, we can always find a step size t_i in (c) above such that

$$F_\varepsilon(x_{i+1}) + \langle x^*, x_{i+1} \rangle < F_\varepsilon(x_i) + \langle x^*, x_i \rangle.$$

This sequence $\{F_\varepsilon(x_i) + \langle x^*, x_i \rangle\}_i$ is by definition bounded below by $E_\varepsilon(x^*)$ and hence convergent.

Next, we set $e_\varepsilon(x^*) = \lim_{i\to\infty}(F_\varepsilon(x_i) + \langle x^*, x_i \rangle)$. By (10) of Theorem 10 (X is reflexive because we assumed X^* uniformly convex)

$$F_\varepsilon(x_i) + \langle x^*, x_i \rangle \geq E_\varepsilon(x^*) = E(x^*) - \varepsilon\phi(x^*).$$

Consequently, $e_\varepsilon(x^*) + \varepsilon\phi(x^*)$ is an upper bound to the (un-regularized) ground-state energy $E(x^*)$. $\qquad\square$

Remark 4. The modification of the Kohn–Sham iteration to include a reduced step size (see (c) in Theorem 12) and also the convergence of the energy quantity $F_\varepsilon + x^*$ were modeled after [6]. These authors mistakenly claim that "the KS algorithm described above is guaranteed to converge". But neither was convergence proven in the usual sense within Banach spaces nor must a converging sequence of potentials lead to the correct x^*_{KS}. Additionally, the work [6] does not use the regularized version of the functionals and thus differentiability (v-representability) cannot be guaranteed. Further investigations of the Kohn–Sham iteration scheme within the framework established here is needed to determine whether a stronger version of convergence can be achieved.

IV. APPLICATIONS

A. Standard density-functional theory

The standard setting of DFT for an N-electron quantum system governed by Coulombic repulsion was pioneered by Lieb [1], but without the tools of Moreau–Yosida regularization or a study of the Kohn–Sham iteration scheme. Adopting the setting to our framework is straightforward.

We now change to standard notation and set $x = \rho$ and $x^* = v$. The Levy–Lieb functional $\tilde{F}$ is defined on the state space of physical densities [1], so-called N-representable densities,

$$\tilde{X} = \{\rho \in L^1(\mathbb{R}^3) \mid \rho \geq 0, \nabla\sqrt{\rho} \in L^2(\mathbb{R}^3), \|\rho\|_1 = N\},$$

which includes all states of finite kinetic energy. Furthermore, $\tilde{X} \subset L^3(\mathbb{R}^3) \cap L^1(\mathbb{R}^3)$ by Sobolev embedding [1].

The functional is then given as the expectation value of the universal part of the standard Hamiltonian,

$$H_\lambda = -\frac{1}{2}\sum_i \nabla_i^2 + \lambda\sum_{i<j}\frac{1}{r_{ij}},$$

with a constrained-search over all wave functions ψ that yield a given density $\rho \in \tilde{X}$,

$$\tilde{F}(\rho) = \inf_{\psi\to\rho}\{\langle\psi, H_1\psi\rangle\}.$$

Note that $\lambda = 1$ corresponds to the interacting system. The functional for the non-interacting ($\lambda = 0$) Kohn–Sham system is similarly given by

$$\tilde{F}^0(\rho) = \inf_{\psi\to\rho}\{\langle\psi, H_0\psi\rangle\}$$

or a restriction of the minimization domain to only those ψ that are Slater determinants.

By the convex conjugate transformations (1) and (2), the functionals E and F are defined on the larger space of densities $L^3(\mathbb{R}^3)\cap L^1(\mathbb{R}^3)$ and the dual space of potentials $L^{3/2}(\mathbb{R}^3)+L^\infty(\mathbb{R}^3)$, which includes singular Coulomb potentials. With $X = L^3\cap L^1$ and $X^* = L^{3/2}+L^\infty$, the convex Lieb functional F is nowhere differentiable but the set of v-representable densities $\mathrm{dom}(\partial F)$ is dense in the set of N-representable densities $\mathrm{dom}(F)$ [1, Theorem 3.14]. The concave ground-state energy E is superdifferentiable at all potentials v that support an electronic ground state and differentiable whenever the ground-state density is nondegenerate. We have

$$-v \in \underline{\partial}F(\rho) \iff \rho \in \overline{\partial}E(v).$$

Kvaal and Helgaker [18] showed that

$$\rho \in \overline{\partial}E(v) \iff E(v) = \mathrm{tr}\,\Gamma_\rho H(v),$$

where $H(v) = H_1 + \sum_i v(r_i)$ is the Hamiltonian with potential v and Γ_ρ is a ground-state density matrix with density ρ. It follows that $\overline{\partial}E(v)$ contains precisely all (ensemble) ground-state densities associated with the potential v, while $\underline{\partial}F(\rho)$ contains all potentials associated with the ground-state density ρ. Clearly, if $H(v)$ does not have a ground state, then $\overline{\partial}E(v) = \emptyset$.

The Hohenberg–Kohn theorem, as the cornerstone of DFT, drops out "for free" in the regularized version from strict monotonicity of $\overline{\partial}E_\varepsilon$, see Section II E. This property follows in turn from E_ε strictly concave. Yet without Moreau–Yosida regularization this strict concavity of E is not at hand and arriving at the usual Hohenberg–Kohn theorem requires a refined analysis of the ground-state density that must not be zero on a set of nonzero measure [19].

To continue, the choice $X = L^3\cap L^1$ ($X^* = L^{3/2}+L^\infty$) does not fit the framework developed here since the L^1-L^∞ pair destroys reflexivity and thus also uniform convexity. A simple solution lies in just widening the density space to $X = L^3(\mathbb{R}^3)$, which includes $\tilde{X}$. The dual space

for potentials is then restricted to $X^* = L^{3/2}(\mathbb{R}^3)$, which is uniformly convex as required by the above theorems. Coulomb potentials on all of $\mathbb{R}^3$ are then ruled out but are still included if the spatial domain is limited to a bounded $\Omega \subset \mathbb{R}^3$.

Conventionally, the Kohn–Sham approach leads to the minimization problem (see the work of [20] and [21] for more technical details)

$$
\begin{aligned}
E(v) &= \inf_\psi \{F(\rho_\psi) + \langle v, \rho_\psi \rangle\} \\
&= \inf_\psi \{T(\psi) + J(\rho_\psi) + F_{\text{xc}}(\rho_\psi) + \langle v, \rho_\psi \rangle\}, \quad (12)
\end{aligned}
$$

where $T(\psi)$ is the kinetic energy, $J(\rho_\psi)$ is the Hartree term, and $F_{\text{xc}}(\rho) = F(\rho) - F^0(\rho) - J(\rho)$ is the exchange-correlation functional. Also, we have above used the notation ρ_ψ to indicate that the density ρ has been computed from ψ, i.e.,

$$
\rho_\psi(x) = \int_{\mathbb{R}^{3(N-1)}} |\psi|^2 dx_2 \cdots dx_N.
$$

Glossing over the fact that $F(\rho)$ is not differentiable, the stationary condition for the above minimization gives the Kohn–Sham equations. When the iterative procedure defined in Theorem 12 converges to the minimum, one obtains the Kohn–Sham potential v_{KS}. Given this potential, it then only requires finding the ground state of a fixed, non-interacting Hamiltonian $H_0 + v_{\text{KS}}$ in order to determine the Kohn–Sham wave function ψ_{KS} that solves the wave function optimization problem in (12) above.

Another related problem is that of finding a Kohn–Sham potential v_{KS} that reproduces the density ρ_{gs} of the ground state of the interacting Hamiltonian $H_1 + v_{\text{ext}}$. This can be done using the optimization problem

$$
F^0(\rho_{\text{gs}}) = \sup_v (E^0(v) - \langle v, \rho_{\text{gs}} \rangle). \quad (13)
$$

When the iterative procedure defined in Theorem 12 converges to the minimum of the energy functional, it in fact solves this problem too *in a way that does not require a priori knowledge of v_{gs}*. Instead, the interacting density ρ_{gs} is specified only implicitly by specifying the external potential v_{ext}.

In summary, although the Kohn–Sham wave function optimization problem (12) and the Lieb optimization problem (13) are distinct, the iterative procedure analyzed above addresses both problems. However, in solving the v-representability problem, we have introduced non-N-representable densities, that is, densities $\rho \notin \text{dom}(F)$. The question then arises regarding the representation of such densities. The non-N-representable densities are obtained in step (b) of Theorem 12 as a supergradient of the regularized non-interacting energy $\rho'_{i+1} \in \overline{\partial} E_\varepsilon^0(v_{i+1})$. Since

$$
\overline{\partial} E_\varepsilon^0(v_{i+1}) = \overline{\partial} E^0(v_{i+1}) - \varepsilon J^{-1}(v_{i+1}),
$$

we may first select an element of $\overline{\partial} E^0(v_{i+1})$ in the usual manner (by solving the non-interacting Schrödinger equation in the regularized potential) and then add the regularization correction $-\varepsilon J^{-1}(v_{i+1})$. See Section VI.B in [7] for further details on how the regularization modifies the Kohn–Sham eigenvalue problem.

B. Current-density-functional theories

In current-density-functional theory (CDFT), both the paramagnetic current density and the total (physical) current density can be used together with the particle density ρ. We refer to [22] and [23] for a discussion of the choice of variables in CDFT. For the specific case of uniform magnetic fields, the current-density degrees of freedom can be reduced into a theory that has been named LDFT [24].

The work of Lieb in DFT [1] was in parts continued into paramagnetic CDFT in [25], where it was proven that each component of the paramagnetic current density is an element of $L^1(\mathbb{R}^3)$. However, since L^1 does not fulfill the requirements presented here, further analysis of function spaces for the paramagnetic formulation is needed. Nonetheless, we conjecture that each component of the paramagnetic current is an element of $L^1 \cap L^q$ for $1 < q < 2$ and we suggest $L^{3/2}$ as a suitable space to choose. We moreover point out that the work in [21] only addressed the problem of v-representability by generalizing the work in [20] to include paramagnetic current densities. The problem of differentiability was not dealt with. The application of the theory outlined here to CDFT formulated with the paramagnetic current density is left for future work and will be based on the above conjecture.

As far as the total (physical) current density is concerned, recent work has established a density-functional theory based on the Maxwell–Schrödinger model [8]. In this theory, the potential space contains pairs $x^* = (v, B)$ of electrostatic scalar potentials and magnetic fields. Holding v fixed, the magnetic self energy plays the role of ϕ above, yielding a ground state energy $E(x^*)$ that is already a Moreau–Yosida regularization with respect to the argument B. The formalism admits construction of a universal density functional $F(x)$ defined for pairs $x = (\rho, \beta)$, where ρ is the electron density and β a type of internal magnetic field that plays the role of an independent variational parameter. Moreover, for any fixed ρ in its domain, the universal functional is differentiable with respect to β. A further regularization with respect to ρ results in functionals $E_\varepsilon(x^*)$ and $F_\varepsilon(x)$ that are within the scope of the above convergence result.

The Maxwell–Schrödinger model is itself an approximation to a more complete description taking into account the quantized nature of the light field that generates the internal magnetic field β [26]. This more complete description is based on the Pauli–Fierz Hamiltonian of non-relativistic quantum electrodynamics [26, 27], which describes the interaction among charged particles (electrons and effective nuclei) by the exchange of pho-

tons, the fundamental gauge bosons of the electromagnetic force [28]. Consequently, the resulting density-functional reformulation is a multi-component theory of fermions and bosons [9] and we have two potentials $x^* = (v, j)$ that act on the respective particle families. Here v is the usual electrostatic scalar potential acting on the electrons and j is an external classical transversal charge current that acts on the photons. Using the standard Maxwell relations, this external current can be directly related to a unique classical magnetic field B [9]. The conjugate pair $x = (\rho, A)$ is then the usual electronic density and the transversal electromagentic vector potential A that is generated by the photons of the coupled matter-photon system. Again, by using the Maxwell relations, the transversal vector potential A is uniquely associated with an internal magnetic field β. The above discussed Moreau–Yosida regularization can then be applied to quantum-electrodynamical DFT, and a rigorous Kohn–Sham iteration scheme based on coupled Maxwell–Pauli–Kohn–Sham equations can be introduced. We note that the presented density-functionalization and general-ization of the Kohn–Sham procedure are applicable in a straightforward manner to other coupled fermion-boson problems. This highlights the applicability of the introduced approach beyond the usual confinement of traditional electronic-structure theory.

[1] E. H. Lieb, Int. J. Quant. Chem. **24**, 243 (1983).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] E. Cancès and C. Le Bris, Int. J. Quant. Chem. **79**, 82 (2000).
[4] E. Cancès, in *Mathematical Models and Methods for ab initio Quantum Chemistry*, edited by M. Defranceschi and C. Le Bris, Lecture Notes in Chemistry, Vol. 74 (Springer, 2000).
[5] E. Cancès, J. Chem. Phys. **144**, 10616 (2001).
[6] L. O. Wagner, E. M. Stoudenmire, K. Burke, and S. R. White, Phys. Rev. Lett. **111**, 093003 (2013).
[7] S. Kvaal, U. Ekström, A. M. Teale, and T. Helgaker, J. Chem. Phys. **140**, 18A518 (2014).
[8] E. I. Tellgren, Phys. Rev. A **97**, 012504 (2018).
[9] M. Ruggenthaler, arxiv:1509.01417v2 (2017).
[10] F. G. Eich, M. Di Ventra, and G. Vignale, J. Phys.: Condens. Matter **29**, 063001 (2016).
[11] K. Pernal and K. J. H. Giesbertz, in *Density-Functional Methods for Excited States*, edited by N. Ferré, M. Filatov, and M. Huix-Rotllant, Topics in Current Chemistry, Vol. 368 (Springer, 2015).
[12] K. J. H. Giesbertz and M. Ruggenthaler, arxiv:1710.08805 (2017).
[13] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces* (4th edition, Springer, 2012).
[14] M. Levy, PNAS **76**, 6062 (1979).
[15] P. E. Lammert, Int. J. Quantum Chem. **107**, 1943 (2007).
[16] O. H. Hiltunen, *Properties of the duality map* (2016) unpublished.
[17] P. Blanchard and E. Brüning, *Mathematical Methods in Physics: Distributions, Hilbert Space Operators, Variational Methods, and Applications in Quantum Physics* (2nd edition, Birkhäuser, 2015).
[18] S. Kvaal and T. Helgaker, J. Chem. Phys. **143**, 184106 (2015).
[19] P. E. Lammert, J. Math. Phys. **59**, 042110 (2018).
[20] N. Hadjisavvas and A. Theophilou, Phys. Rev. A **30**, 2183 (1984).
[21] A. Laestadius, J. Math. Chem. **52**, 2581 (2014).
[22] E. I. Tellgren, S. Kvaal, E. Sagvolden, U. Ekströ m, A. M. Teale and T. Helgaker, Phys. Rev. A **86**, 062506 (2012).
[23] A. Laestadius and M. Benedicks, Int. J. Quant. Chem. **114**, 782 (2014).
[24] E. I. Tellgren, A. Laestadius, T. Helgaker, S. Kvaal and A. M. Teale, J. Chem. Phys. **148**, 024101 (2018).
[25] A. Laestadius, Int. J. Quant. Chem. **114**, 1445 (2014).
[26] M. Ruggenthaler, N. Tancogne-Dejean, J. Flick, H. Appel, A. Rubio, Nature Reviews Chemistry **2**, 0118 (2018).
[27] H. Spohn, *Dynamics of charged particles and their radiation field* (Cambridge University Press, 2004).
[28] L. H. Ryder, *Quantum field theory* (Cambridge University Press, 1996).