HAL
open science

Multi-center decomposition of molecular densities: a mathematical perspective

Robert Benda, Eric Cancès, Virginie Ehrlacher, Benjamin Stamm

To cite this version:

Robert Benda, Eric Cancès, Virginie Ehrlacher, Benjamin Stamm. Multi-center decomposition of molecular densities: a mathematical perspective. The Journal of Chemical Physics, 2022, 156, pp.164107. 10.1063/5.0076630 . hal-03338862

HAL Id: hal-03338862

https://hal.science/hal-03338862

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multi-center decomposition of molecular densities: a mathematical perspective

Robert Benda[1,2], Eric Cancès[1], Virginie Ehrlacher[1], and Benjamin Stamm[3]

[1]CERMICS, Ecole des Ponts and Inria Paris, 6 & 8 avenue Blaise Pascal, 77455 Marne-la-Vallée, France
[2]LPICM, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France
[3]ACoM, Department of Mathematics, RWTH Aachen University, Schinkelstrasse 2, D-52062 Aachen, Germany

September 8, 2021

Abstract

The aim of this paper is to analyze from a mathematical perspective some existing schemes to partition a molecular density into several atomic contributions, with a specific focus on Iterative Stockholder Atom (ISA) methods. We provide a unified mathematical framework to describe the latter family of methods and propose a new scheme, named L-ISA (for linear approximation of ISA). We prove several important mathematical properties of the ISA and L-ISA minimization problems and show that the so-called ISA algorithms can be viewed as alternating minimization schemes, which in turn enables us to obtain new convergence results for these numerical methods. Specific mathematical properties of the ISA decomposition for diatomic systems are also presented. We also review the basis-space oriented Distributed Multipole Analysis method, the mathematical formulation of which is also clarified. Different schemes are numerically compared on different molecules and we discuss the advantages and drawbacks of each approach.

Contents

1 Introduction **2**

2 AIM methods based on objective functionals **4**
2.1 Set of admissible AIM decompositions . . . 5
2.2 Main ingredients of an AIM decomposition method . . . 6
2.3 Proatom densities . . . 6
2.4 Objective functionals . . . 8
2.5 Kullback-Leibler entropy methods . . . 9
2.5.1 The Hirschfeld method . . . 9
2.5.2 AIM iterative methods based on the Kullback-Leibler divergence . . . 10
2.5.3 Hirschfeld-I iterations . . . 12
2.5.4 ISA iterations . . . 13
2.5.5 Linear approximation of the ISA (L-ISA) method . . . 14
2.5.6 Nonlinear approximation of the ISA (NL-ISA) . . . 16
2.6 GISA iterations . . . 16
2.7 MB-ISA iterations . . . 18

3 Mathematical analysis of AIM decomposition methods **19**

4 Distributed Multipole Analysis **22**
4.1 Multipole moments with respect to natural centers . . . 23
4.2 Redistribution formula . . . 23
4.3 Multipolar expansion of the electrostatic potential . . . 25

5 Numerical results **26**
5.1 Implementational details . . . 26
5.2 Comparison of the different ISA schemes . . . 29
5.3 Dissociation of diatomic molecules . . . 33
5.4 Basis set convergence of DMA and ISA . . . 35

6 Proofs **36**
6.1 Proof of Lemma 2 . . . 37
6.2 Proof of Lemma 5 . . . 38
6.3 L-ISA counterpart of Lemma 5 . . . 38
6.4 Topological properties of the set of admissible AIM decompositions . . . 39
6.5 Proof of Theorem 6 . . . 40
6.6 Proof of Theorem 7 . . . 43
6.7 Proof of Proposition 8 . . . 46
6.8 Proof of Proposition 9 . . . 48

1 Introduction

A successful study of intermolecular interactions is closely interwoven with the quality of the representation of the electronic density and subsequently of the electrostatic potential (ESP) around molecules. When a continuous representation of the density, such as in Density Functional Theory (DFT) [26, 28] or Wave-Function Theory (WFT) [27], cannot be used because of a too high computational burden, the density has to be summarized by a finite number of parameters thanks to *localization schemes*. Classical force fields, such as GROMOS [49], CHARMM [35] or AMBER [42, 13], among others, use such a representation, with (fixed) so-called partial charges assigned to all atomic sites, to represent the electronic density in a simplified way and account for electrostatic interactions with neighboring molecules. Molecular dynamics (MD) simulations based on these force fields are commonly used to gain insight into the structural or energetic properties of biologic or organic compounds. The physical relevance of MD studies heavily relies on their quality, and thus on the specific localization scheme used to compute them. The purpose of the present paper is precisely to review several localization schemes from a mathematical perspective, providing a sound mathematical basis to the latter. Let us first review the existing localization schemes that have been used to derive atomic partial charges or atomic multipole moments in the context of force field parameterization.

First, a popular method to derive partial charges is to fit the ESP generated by the set of atomic point charges to the exact quantum ESP (derived from DFT or WFT calculations) at a given geometry [51], possibly imposing additional constraints to lower the conformational dependence of the obtained optimal partial charges, such as in the RESP method [6]. Adding constraints to reproduce the total dipole moments at different conformations is also possible [47]. So is to fit the ESP generated by the atomic charges to the approximate ESP generated by a series of distributed multipole moments [52, 54], as implemented *e.g.* in the Mulfit code [15]. In both cases, the partial charges are chosen as those leading to the smallest error on the ESP

for a suitable norm. This family of ESP-derived charges methods amounts to find indirectly the partial charges from the quantum calculation, using the electrostatic potential as a proxy – and possibly the molecular dipole or quadrupole moments – as relevant target. However, they are associated to several drawbacks, such as conformational-dependent charges – although strategies to mitigate this undesired feature have been designed [47] – and lack of transferability. Chipot *et al.* have for instance shown that ESP derived partial charges are not transferable from one to another similar molecule [12]. In other words, the partial charge assigned to a carbon atom in a given specific local chemical environment of a given molecule cannot be used as partial charge for a carbon atom in a similar local chemical environment of another molecule. It is also possible to derive atomic partial charges directly from a quantum calculation (namely, using the coefficients of the converged density matrix) using *e.g.* Mulliken [38] or Löwdin [33] population analysis, although these methods are not commonly used for partial charges parameterization – or, if they are, charges are readjusted afterwards by fitting to the ESP, to minimize the deviation between the reference QM ESP and the atomic charges-generated ESP. Indeed, Mulliken and Löwdin schemes adopt a basis space approach, which results in significant basis-set dependency of partial charges and lack of convergence with increasing basis set size.

Real-space approaches, using a Voronoi-like partitioning of the space, such as the Bader Atom In Molecules [5], the Becke space-decomposition (fuzzy Voronoi cells) approach [7], or stockholder partitionning schemes originating from the Hirshfeld method [25] (see below), have also been used to derive partial charges, and do not suffer as much from basis set dependence – but may suffer other drawbacks [23].

Finally, thanks to the recent availability of large databases of molecules already parameterized for classical force fields, modern machine-learning inspired methods have also been used to estimate partial charges, *e.g.* formulating the charge assignment problem as a multiple-choice knapsack problem [14]. Yet, although these methods allow to avoid explicit charge calculations, they may end up with atomic charges suffering from the same drawbacks as those of the database on which they have been fitted.

To correct for the too approximate nature of the ESP represented by atomic partial charges only [12] (among other drawbacks), higher order moments of the charge distribution have been introduced within localization schemes, and consequently in force fields used for MD simulations. At their early times, multipole moments of the electronic density were introduced as a generalization of Mulliken population analysis, to analyse and interpret molecular wave-functions with a larger number of local descriptors [52, 54, 60]. Since the advent of modern polarizable force fields, such as AMOEBA [44, 45, 46, 62], these local descriptors have also been used successfully to improve the description of biomolecular and organic molecules, and of their mutual interactions, in particular thanks to an improved treatment of the water solvent. Local multipole moments – which are used *e.g.* up to quadrupole moments in AMOEBA force field – allow to capture the local anisotropy of the charge density (contrary to atomic charges alone) and are thus a key ingredient of polarizable force fields.

In a similar manner as for partial charges, local multipole moments can be derived either *indirectly* (by fitting the multipoles-derived ESP to the quantum ESP) or *directly* (*e.g.* using the density matrix corresponding to a certain basis expansion, or the values of the density on a grid) from the result of a quantum calculation. To the former rationale belongs the fitting method implemented *e.g.* in the Mulfit code [16, 61], which allows to derive atomic charges [61] or multipoles [16] (up to user-specified orders, for every atom) by fitting on the ESP generated by higher order local multipoles – typically obtained previously from distributed multipole analysis (DMA) [52, 54]. To the latter family of methods belong Stone's DMA [52, 54], performing a redistribution of the density matrix coefficients and elementary local multipole moments to final expansion sites (*e.g.* all the atoms), and real-space partitioning schemes in the continuity

of Bader concepts of atoms in molecules [4]. The family of Hirshfeld-like partitionning schemes [24, 36, 3, 41], namely its iterative version Hirshfeld-I [10, 11], the iterative stockholder approach (ISA) [30, 31] and its variants GISA [57], MB-ISA [59], and more recent basis-space implementation (BS-ISA) [37], belong to this family of real-space oriented methods. This family of methods yields multipole moments which have been claimed to be rather transferable [23]. Note that in principle, real-space partitioning schemes end up with local multipole moments located at atomic sites only (to enable force computations in force field applications), although this is not a requirement. The DMA method can also naturally allocate contributions to atoms and to other non-atomic sites, such as bond centers.

In this article, we review and clarify several existing schemes to partition a molecular density into several (*e.g.* atomic) contributions or to derive directly distributed multipole moments. In Section 2, we first focus on the real-space oriented methods of the Hirshfeld / Iterative Stockholder Atom (ISA) family. We provide a unified mathematical framework to describe the existing ISA methods, through the formulation (21). This formulation is based on information theory, a link that was first introduced by Nalewajski and Parr (see [39]) and then subsequently used in the works of Ayers, Bultinck, Heidar-Zadeh and co-workers (see [10, 11, 23] and references therein), but is new to our knowledge. It is also worthwhile to stress the mathematical properties established in [3]. Driven by formulation (21), we propose a new AIM scheme, named L-ISA (linear approximation of ISA), and presented in Section 2.5.5. We also formulate the Minimal-Basis ISA (MB-ISA) method as a minimization of the Kullback-Leibler entropy between atomic-*shell* densities and the corresponding pro-atomic-*shell* densities, under equality constraints involving zero and first order moments of the atomic densities (see section 2.7). We then focus in Section 3 on the mathematical analysis of the ISA and L-ISA methods. We first prove the existence and uniqueness of the solutions to the ISA and L-ISA optimization problems (Theorem 6). While uniqueness easily follows from the strict convexity properties of the Kullback-Leibler entropy, establishing the existence of a minimizer is more difficult due to possible loss of compactness. We then prove that the ISA and L-ISA algorithms can be interpreted as alternating minimization methods, that the associated entropy is a Lyapunov function of the algorithm, and that the L^2 norm between two successive iterates converges to 0. We also prove that, in the case of the L-ISA method, the L-ISA algorithm converges toward the unique minimizer of the L-ISA optimization problem. All these results are collected in Theorem 7. We finally prove specific mathematical properties of ISA decomposition for diatomic systems in Propositions 8 and 9. In Section 4, we review a popular basis-space oriented method, the Distributed Multipole Analysis (DMA), whose mathematical formulation is also clarified. Finally, we present in Section 5 some numerical results (with a focus on diatomic systems) of the different ISA schemes, namely GISA, L-ISA (section 2.5.5), MB-ISA and (historical) ISA (section 2.5.4), which have been implemented numerically. The DMA method, which has also been implemented in a more modern and modular form, enabling for more redistribution strategies to be explored (section 4), is also compared to the ISA method in terms of convergence of local multipole moments with increasing basis set size, for some test systems.

2 AIM methods based on objective functionals

We denote by

$$X := \left\{ f \in L^1(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3) \;\middle|\; \lim_{|\mathbf{r}|\to\infty} f(\mathbf{r}) = 0,\ \int_{\mathbb{R}^3} |\mathbf{r}||f(\mathbf{r})|\, d\mathbf{r} < \infty \right\}$$

the Banach space of bounded, integrable, real-valued functions on the physical space $\mathbb{R}^3$ with finite first moments, and vanishing at infinity. In the sequel, we consider a given nonnegative

density $\rho \in X$ such that

$$N := \int_{\mathbb{R}^3} \rho(\mathbf{r})\, d\mathbf{r} > 0,$$

and a collection $\mathbf{R} = (\mathbf{R}_a)_{1\le a\le M} \in (\mathbb{R}^3)^M$ of M points in $\mathbb{R}^3$. Ground and excited state electronic densities of molecules and clusters are known to be continuous, positive, functions on $\mathbb{R}^3$ decaying exponentially fast at infinity (see e.g. [18, Theorem 1.3]), and are therefore elements of X.

2.1 Set of admissible AIM decompositions

The goal is to decompose ρ as a sum of translated nonnegative densities $\rho_a^{\rm opt} \in X$, i.e.

$$\rho(\mathbf{r}) = \sum_{a=1}^{M} \rho_a^{\rm opt}(\mathbf{r} - \mathbf{R}_a), \tag{1}$$

each $\rho_a^{\rm opt}$ being localized around the origin and optimized in some sense. In quantum chemistry, the $\mathbf{R}_a$'s are most often the positions of the M nuclei of a molecular system containing $N \in \mathbb{N}^*$ electrons, and ρ an approximation of its ground-state (or k-th excited-state) electronic density obtained by a given electronic structure calculation method in a given basis set. We will then denote by $z_a \in \mathbb{N}^*$ the charge of nucleus a for $1 \le a \le M$. However, in some applications, it can be useful to choose expansion centers away from the nuclei, for instance on a chemical bond, or at the center of mass of a functional group. We will therefore call the points $(\mathbf{R}_a)_{1\le a\le M}$ the expansion centers (rather than the atomic positions); for convenience, we will however use the standard terminology *atoms-in-molecules* (AIM) to refer to the family of functions $(\rho_a^{\rm opt})_{1\le a\le M}$.

For later purposes, we introduce the atomic charge linear map $\boldsymbol{\mathcal{N}} : X^M \to \mathbb{R}^M$ defined by

$$\forall \boldsymbol{\rho} := (\rho_a)_{1\le a\le M} \in X^M, \quad [\boldsymbol{\mathcal{N}}(\boldsymbol{\rho})]_a = \int_{\mathbb{R}^3} \rho_a(\mathbf{r})\, d\mathbf{r},\ \forall 1 \le a \le M. \tag{2}$$

Remark 1. *We use the decomposition (1) instead of the (equivalent) usual decomposition* $\rho(\mathbf{r}) = \sum_{a=1}^{M} \rho_a^{\rm opt}(\mathbf{r})$ *because the former is more convenient for the analysis of the dependence of the AIM decomposition on the atomic positions. For instance, it is expected that the* $\rho_a^{\rm opt}$ *(centered at the origin) in (1) converge to the ground (or excited) state density of the isolated atom or ion in the dissociation limit.*

The $\rho_a^{\rm opt}$'s can be used as such, or transformed into a collection of descriptors, typically their charges, and their first (dipolar) and second-order moments:

$$q_a := \int_{\mathbb{R}^3} \rho_a^{\rm opt}(\mathbf{r})\, d\mathbf{r}, \quad \mathbf{p}_a := \int_{\mathbb{R}^3} \mathbf{r}\rho_a^{\rm opt}(\mathbf{r})\, d\mathbf{r}, \quad \mathbf{Q}_a := \int_{\mathbb{R}^3} \mathbf{r}\otimes\mathbf{r}\, \rho_a^{\rm opt}(\mathbf{r})\, d\mathbf{r}.$$

We denote by $X_+ := \{f \in X \mid f \ge 0 \text{ a.e.}\}$ the convex cone of bounded integrable nonnegative densities vanishing at infinity and introduce

$$\boxed{\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} := \left\{ \boldsymbol{\rho} = (\rho_a)_{1\le a\le M} \in X_+^M \;\middle|\; \sum_{a=1}^{M} \rho_a(\,\cdot\, - \mathbf{R}_a) = \rho \right\}}, \tag{3}$$

the set of admissible AIM decompositions of the density ρ with respect to the expansion centers $\mathbf{R} = (\mathbf{R}_a)_{1\le a\le M} \in (\mathbb{R}^3)^M$.

The set $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ is non-empty and convex. It has also some nice topological properties which will allow us to prove the existence of an optimal decomposition for the usual AIM decomposition methods (see the mathematical results in Section 3 and their proofs in Section 6).

2.2 Main ingredients of an AIM decomposition method

Let us denote by $X^r \subset X$ (respectively $X_+^r \subset X_+$) the subset of functions of X (respectively X_+) that are radially symmetric.

An objective-functional-based AIM decomposition method consists in seeking an optimal decomposition $\boldsymbol{\rho}^{\mathrm{opt}} := (\rho_a^{\mathrm{opt}})_{1\le a\le M} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$, for a specific criterion. It is characterized by two main ingredients:

(i) a collection of sets of (radially symmetric) proatom densities $(\mathcal{K}_z^0)_{z\in\mathbb{N}^*}$, so that for each $z \in \mathbb{N}^*$, $\mathcal{K}_z^0 \subset X_+^r$ contains reasonable approximations of the ground state density of an isolated atom with nuclear charge z. The set of proatom densities is then used to build a set of admissible proatoms-in-molecule (or promolecule) densities

$$\boxed{\boldsymbol{\mathcal{K}}^0 := \mathcal{K}_{z_1}^0 \times \cdots \times \mathcal{K}_{z_M}^0 = \left\{\boldsymbol{\rho}^0 := (\rho_a^0)_{1\le a\le M},\ \rho_a^0 \in \mathcal{K}_{z_a}^0\right\} \subset (X_+^r)^M} \tag{4}$$

(ii) an objective functional $\mathcal{E} : X_+^M \times (X_+^r)^M \to \mathbb{R} \cup \{+\infty\}$, usually in the form of a relative entropy.

The set $\mathcal{K}_z^0$ may, or not, depend explicitly on z. It may have the structure of a vector space but does not need to in general. For all $1 \le a \le M$, it is expected that, ρ_a^{opt} should be close in some sense to one of the elements of $\mathcal{K}_{z_a}^0$, but let us emphasize that, in general, ρ_a^{opt} will not be a radially symmetric function, hence will not belong to $\mathcal{K}_{z_a}^0$.

The quantity $\mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0)$ measures the discrepancy between a trial AIM decomposition $\boldsymbol{\rho} := (\rho_a)_{1\le a\le M} \in X_+^M$ and a trial promolecule density $\boldsymbol{\rho}^0 := (\rho_a^0)_{1\le a\le M} \in \boldsymbol{\mathcal{K}}^0$. Most objective-functional-based AIM decomposition methods (including ISA, Hirschfeld, Hirschfeld-I, L-ISA, NL-ISA) consist in finding $\left(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}}\right) \in \boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}}$ solution to the minimization problem

$$\boxed{\left(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}}\right) \in \underset{(\boldsymbol{\rho},\boldsymbol{\rho}^0)\in\boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}}}{\operatorname{argmin}} \ \mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0)} \tag{5}$$

where $\boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}}$ is a well-chosen subset of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0$.

In Sections 2.3 and 2.4, we detail various possible choices of sets of proatom densities $\mathcal{K}_z^0$ and objective functionals which are encountered in the literature. The aim of Section 2.5 is to present in more details the methods based on the Kullback-Leibler divergence, including Hirschfeld, Hirschfeld-I, ISA, L-ISA and NL-ISA. The GISA and MB-ISA methods cannot be expressed under the form (5); their mathematical structures are presented in Sections 2.6 and 2.7 respectively.

2.3 Proatom densities

As anticipated above, we now discuss several choices for the proatom densities $\mathcal{K}_z^0$.

$\boxed{1}$ In the original *Hirschfeld method* [24], the set $\mathcal{K}_z^0$ is a singleton and only contains the ground state density of the neutral atom

$$\mathcal{K}_{z,\mathrm{Hirschfeld}}^0 = \left\{\rho_{z,z}^0\right\},$$

where $\rho_{z,n}^0$ denotes (an approximation of) the radially symmetric ground state density of the atomic system consisting of a single nucleus of charge z and n electrons[1], being understood

[1] If the ground state is degenerate and the radial symmetry broken, $\rho_{z,n}^0$ is chosen equal to the radially symmetric mixed-state ground-state density obtained by averaging the pure-state ground-state densities (with respect to the Haar measure of the rotation group SO(3)).

that $\rho^0_{z,n} = \rho^0_{z,N_z^{\max}}$ for n greater than $N_z^{\max}$, the maximum number of electrons that a nucleus of charge z can bind. However, choosing for ρ^0_a electronic densities of *neutral* atoms does not seem appropriate for strongly polarized molecules where significant charge transfers occur. This and other drawbacks are detailed in [11], such as the fact that the Hirshfeld method is limited to neutral molecules, and that Hirshfeld atomic charges are on average too small to accurately describe molecular polarization. A desirable feature of an AIM method is its ability to detect charge transfer in an automatic way, without requiring *a priori* knowledge from the user on the chemical system under consideration.

$\boxed{2}$ In the so-called *Hirschfeld-I method* [11] (where I stands for iterative), this is achieved to some point by the following choice of proatom densities which contains not only the neutral ground-state atomic density, but also the ground-state densities of the ionized forms of the atom, as well as those of the mixed states corresponding to fractional (in the sense of non-integer) numbers of electrons:

$$\mathcal{K}^0_{z,\mathrm{Hirschfeld-I}} = \left\{\rho^0_{z,n},\, n \in \mathbb{R}_+\right\}$$

with

$$\forall n \in \mathbb{R}_+, \quad \rho^0_{z,n}(\mathbf{r}) := (\lceil n \rceil - n)\, \rho^0_{z,\lfloor n \rfloor}(\mathbf{r}) + (n - \lfloor n \rfloor)\, \rho^0_{z,\lceil n \rceil}(\mathbf{r}),$$

where for $x \in \mathbb{R}$, $\lfloor x \rfloor$ and $\lceil x \rceil$ respectively denote the largest integer smaller than x, and the smallest integer larger than x.

$\boxed{3}$ In the *iterative stockholder approach (ISA)* [30, 31], this set is independent of the chemical element and is chosen as the whole convex cone of bounded integrable nonnegative radially symmetric functions vanishing at infinity:

$$\mathcal{K}^0_{\mathrm{ISA}} = X^r_+.$$

$\boxed{4}$ In the *finite-dimensional linear approximations of ISA (L-ISA)*, the set $\mathcal{K}^0_z$ is chosen to be a non-empty closed convex subset of X^r_+. More precisely,

$$\mathcal{K}^0_{z,\mathrm{L-ISA}} = \left\{\rho^0(\mathbf{r}) = \sum_{k=1}^{m_z} c_k\, g_{z,k}(\mathbf{r}),\ c_k \in \mathbb{R}_+\right\}, \tag{6}$$

where the $g_{z,k}$'s, $1 \leq k \leq m_z$ are z-dependent given linearly-independent, positive, L^1-normalized functions of $\mathcal{K}^0_{\mathrm{ISA}} = X^r_+$.

$\boxed{5}$ In the *finite-dimensional nonlinear approximations of ISA (NL-ISA)*,

$$\mathcal{K}^0_z = \mathcal{K}^0_{z,\mathrm{NL-ISA}} = \left\{\rho^0(\mathbf{r}) = \sum_{k=1}^{m_z} c_k\, g_{z,k,\alpha_k}(\mathbf{r}),\ c_k \in \mathbb{R}_+,\ \alpha_k > 0\right\}, \tag{7}$$

where the g_{z,k,α_k}'s are L^1-normalized positive functions of $\mathcal{K}^0_{\mathrm{ISA}} = X^r_+$ depending on a parameter α_k in a non-affine manner and which needs to be optimized. For the sake of simplicity, here, we assume that the parameters α_k are positive real numbers, but α_k could be a vector subjected to equality and/or inequality constraints.

$\boxed{6}$ In *GISA* [57], the set $\mathcal{K}_z^0$ is defined as the finite-dimensional convex cone

$$\mathcal{K}_{z,\mathrm{GISA}}^0 = \left\{ \rho_z^0(\mathbf{r}) = \sum_{k=1}^{m_z} c_k\, \zeta_{\alpha_{z,k}}(\mathbf{r}),\ c_k \in \mathbb{R}_+ \right\} \quad \text{where} \quad \zeta_\alpha(\mathbf{r}) = \left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha|\mathbf{r}|^2}, \tag{8}$$

and where $m_z \in \mathbb{N}^*$ and $\alpha_{z,k} > 0$ are fixed empirical parameters.

2.4 Objective functionals

As mentioned above, the aim of the quantity $\mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0)$ is to measure the discrepancy (in some sense) between a trial AIM decomposition $\boldsymbol{\rho} := (\rho_a)_{1\leq a\leq M} \in X_+^M$ and a trial promolecule density $\boldsymbol{\rho}^0 := (\rho_a^0)_{1\leq a\leq M} \in \boldsymbol{\mathcal{K}}^0$.

Most of the popular information-theory-based AIM methods (Hirschfeld, Hirschfeld-I, ISA, L-ISA, NL-ISA – see below for further details) make use of the relative entropy

$$\mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0) := S(\boldsymbol{\rho}|\boldsymbol{\rho}^0), \tag{9}$$

where

$$S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) := \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a(\mathbf{r}) \log\left(\frac{\rho_a(\mathbf{r})}{\rho_a^0(\mathbf{r})}\right) d\mathbf{r} = \sum_{a=1}^{M} s_{KL}(\rho_a|\rho_a^0), \tag{10}$$

constructed from the Kullback-Leibler divergence commonly used in information theory:

$$\forall f, g \in X_+, \quad s_{\mathrm{KL}}(f|g) := \int_{\mathbb{R}^3} f(\mathbf{r}) \log\left(\frac{f(\mathbf{r})}{g(\mathbf{r})}\right) d\mathbf{r}. \tag{11}$$

In order to make the above definition for L^1 functions consistent with the general definition of the Kullback-Leibler divergence for bounded positive measures, the following conventions must be used:

$$0 \log\left(\frac{0}{0}\right) = 0 \quad \text{and for all } p > 0, \quad 0 \log\left(\frac{0}{p}\right) = 0, \quad p \log\left(\frac{p}{0}\right) = +\infty. \tag{12}$$

Note however that alternative methods have been proposed [22], where the objective functional $\mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0)$ is chosen as

$$\mathcal{E}(\boldsymbol{\rho}, \boldsymbol{\rho}^0) = \sum_{a=1}^{M} s(\rho_a|\rho_a^0),$$

where $s(\cdot|\cdot)$ is either the Hellinger distance

$$s_{\mathrm{Hel}}(f|g) := \int_{\mathbb{R}^3} \left(\sqrt{f(\mathbf{r})} - \sqrt{g(\mathbf{r})}\right)^2 d\mathbf{r} = \int_{\mathbb{R}^3} f(\mathbf{r}) \left(\sqrt{\frac{g(\mathbf{r})}{f(\mathbf{r})}} - 1\right)^2, \tag{13}$$

or a more general divergence of the form

$$s_\phi(f|g) := \int_{\mathbb{R}^3} f(\mathbf{r}) \phi\left(\frac{g(\mathbf{r})}{f(\mathbf{r})}\right), \tag{14}$$

where $\phi : \mathbb{R}_+ \to \mathbb{R} \cap \{+\infty\}$ is a strictly convex function, smooth on $\mathbb{R}_+^*$ and such that $\phi(1) = 0$.

The GISA method [57] makes use of both the relative entropy $S(\boldsymbol{\rho}|\boldsymbol{\rho}^0)$ and the L^2-distance

$$d_{L^2}(\boldsymbol{\rho}, \boldsymbol{\rho}^0) := \left(\sum_{a=1}^{M} \|\rho_a - \rho_a^0\|_{L^2}^2\right)^{1/2},$$

in a more complex fashion as will be summarized below in Section 2.6.

Lastly, the Minimal Basis Iterative Stockholder algorithm (MB-ISA) [59] is in fact not based on a direct atomic decomposition on ρ. It rather provides an *atomic-shell* decomposition of ρ on *proatomic-shell* densities, from which AIM densities can be derived. As will be seen in Section 2.7, it nevertheless nicely fits in the unified mathematical framework described in this article.

2.5 Kullback-Leibler entropy methods

In this section, we more specifically focus on the AIM decomposition methods the objective functional of which is based on the Kullback-Leibler divergence $S(\cdot|\cdot)$. Let us first recall the well-known formula (16) below, a rigorous proof of which is provided in Section 6.

Lemma 2. *Let* $\boldsymbol{\rho}^0 := (\rho_a^0)_{1\leq a\leq M} \in (X_+^r)^M$ *and* $\rho \in X_+$ *be such that*

$$s_{\mathrm{KL}}(\rho|\rho^0) < +\infty \quad \textit{where} \quad \rho^0(\mathbf{r}) := \sum_{a=1}^{M} \rho_a^0(\mathbf{r}-\mathbf{R}_a). \tag{15}$$

Then, the constrained minimization problem

$$\inf_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0)$$

admits an unique minimizer $\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0) := ([\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0)]_a)_{1\leq a\leq M}$*, given by*

$$\forall 1\leq a\leq M, \quad [\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0)]_a(\boldsymbol{r}) = \left| \begin{array}{ll} \frac{\rho_a^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r}+\mathbf{R}_a)}\rho(\boldsymbol{r}+\mathbf{R}_a) & \textit{if } \rho_a^0(\boldsymbol{r})>0, \\ 0 & \textit{if } \rho_a^0(\boldsymbol{r})=0, \end{array} \right. \tag{16}$$

and it holds that

$$\min_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) = s_{\mathrm{KL}}(\rho|\rho^0). \tag{17}$$

Observe that if $\rho_a^0(\boldsymbol{r}) > 0$, then $\rho^0(\boldsymbol{r}+\mathbf{R}_a) \geq \rho_a^0(\boldsymbol{r}) > 0$, while if $\rho_a^0(\boldsymbol{r}) = 0$, then we can have either $\rho^0(\boldsymbol{r}+\mathbf{R}_a) > \rho_a^0(\boldsymbol{r})$, in which case $\frac{\rho_a^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r}+\mathbf{R}_a)}\rho(\boldsymbol{r}+\mathbf{R}_a)$ is well-defined and equal to 0, or $\rho^0(\boldsymbol{r}+\mathbf{R}_a) = \rho_a^0(\boldsymbol{r}) = 0$, in which case $\frac{\rho_a^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r}+\mathbf{R}_a)}\rho(\boldsymbol{r}+\mathbf{R}_a)$ is not well-defined.

Remark 3. *In the case when* $\mathcal{E}(\boldsymbol{\rho},\boldsymbol{\rho}^0) = S(\boldsymbol{\rho}|\boldsymbol{\rho}^0)$ *and* $\boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}} = \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}\times\boldsymbol{\mathcal{K}}^0$*, we infer from Lemma 2 that the minimization problem in (5) is equivalent to*

$$\inf_{\boldsymbol{\rho}^0:=(\rho_a^0)_{1\leq a\leq M}\in\boldsymbol{\mathcal{K}}^0} s_{\mathrm{KL}}\left(\rho \,\middle|\, \sum_{a=1}^{M}\rho_a^0(\mathbf{r}-\mathbf{R}_a)\right). \tag{18}$$

Although this problem seems natural, it is not the one solved in most AIM methods, probably because it is computationally challenging in general. Only the basic Hirschfeld method fits into this framework, as will be seen in the next section, but in this case, (18) is trivial since $\boldsymbol{\mathcal{K}}^0_{\mathrm{Hirschfeld}} = \{\boldsymbol{\rho}^0_{\mathrm{Hirschfeld}}\}$ *contains only one element.*

2.5.1 The Hirschfeld method

The Hirshfeld method consists in solving

$$\inf_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0_{\mathrm{Hirschfeld}}), \tag{19}$$

where $\boldsymbol{\rho}^0_{\text{Hirschfeld}} = (\rho^0_{z_a,z_a})_{1\le a\le M}$ is given. As already mentioned in Remark 3, this problem fits into the general framework (5) using the objective functional $\mathcal{E}$ defined in (9) and $\boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}} = \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0_{\text{Hirschfeld}}$.

The atomic densities $\rho^0_{z_a,z_a}$ are positive everywhere and decay as exponentials at infinity; in addition, exact molecular densities and their usual approximations decay (at least) exponentially fast at infinity. The assumptions in Lemma 2 are thus satisfied, which implies that the basic Hirschfeld problem (19) is well-posed and has an explicit solution:

$$\rho_a^{\text{opt,Hirschfeld}}(\mathbf{r}) = [\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0_{\text{Hirschfeld}})]_a(\mathbf{r}) = \frac{\rho^0_{z_a,z_a}(\mathbf{r})}{\displaystyle\sum_{b=1}^{N} \rho^0_{z_b,z_b}(\mathbf{r}-\mathbf{R}_b+\mathbf{R}_a)}\rho(\mathbf{R}_a+\mathbf{r}). \tag{20}$$

2.5.2 AIM iterative methods based on the Kullback-Leibler divergence

The ISA, Hirschfeld-I, L-ISA and NL-ISA methods can be formulated using the following unified formalism: the optimal AIM decomposition is obtained by solving

$$\boxed{\inf_{(\boldsymbol{\rho},\boldsymbol{\rho}^0)\in\boldsymbol{\mathcal{C}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0)} \tag{21}$$

where

$$\boxed{\boldsymbol{\mathcal{C}}_{\rho,\mathbf{R}} = \left\{(\boldsymbol{\rho},\boldsymbol{\rho}^0)\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}\times\boldsymbol{\mathcal{K}}^0 \mid \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)\right\}}$$

with $\boldsymbol{\mathcal{N}}$ being the atomic charge map defined in (2). Note that the variation that determines the different methods (ISA, Hirschfeld-I, L-ISA, NL-ISA) lies in the definition of $\boldsymbol{\mathcal{K}}^0$ defined from the $\mathcal{K}^0_z$'s given in Section 2.3, i.e. $\mathcal{K}^0_z = \mathcal{K}^0_{z,\text{Hirschfeld-I}}$, $\mathcal{K}^0_z = \mathcal{K}^0_{\text{ISA}}$, $\mathcal{K}^0_z = \mathcal{K}^0_{z,\text{L-ISA}}$ or $\mathcal{K}^0_z = \mathcal{K}^0_{z,\text{NL-ISA}}$. These methods hence fit into the general framework (5) using the objective functional $\mathcal{E}$ defined in (9) and $\boldsymbol{\mathcal{D}}_{\rho,\mathbf{R}} = \boldsymbol{\mathcal{C}}_{\rho,\mathbf{R}}$ with corresponding $\boldsymbol{\mathcal{K}}^0$. The solution to (21) is not explicit and must be computed numerically, by an iterative algorithm. The constraints $\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)$ request that for all $1 \le a \le M$, ρ_a and ρ^0_a have the same charge, i.e. $\int_{\mathbb{R}^3}\rho_a(\mathbf{r})\,d\mathbf{r} = \int_{\mathbb{R}^3}\rho^0_a(\mathbf{r})\,d\mathbf{r}$, which is not the case for the optimal AIM obtained with the basic (non-iterative) Hirschfeld method.

Remark 4. *For this approach to make sense, the least one can ask is that*

$$\text{there exists } (\boldsymbol{\rho},\boldsymbol{\rho}^0)\in\boldsymbol{\mathcal{C}}_{\rho,\mathbf{R}} \text{ such that } S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) < +\infty. \tag{22}$$

This condition guarantees that the infimum in (21) is finite. Note that it is not satisfied, even for physical ground-state electronic densities ρ, if e.g. the $\mathcal{K}^0_z$'s only contain compactly supported functions. For (22) to be satisfied, the sets $\mathcal{K}^0_z$ must contain functions which do not decay extremely fast at infinity. This is always the case for ISA for which the sets $\mathcal{K}^0_{\text{ISA}} = X^r_+$' contains the function $e^{-|\mathbf{r}|}$, so that taking $\rho_a(\mathbf{r}) = M^{-1}\rho(\mathbf{r}+\mathbf{R}_a)$ and $\rho^0_a(\mathbf{r}) = \frac{N}{8\pi M}e^{-|\mathbf{r}|}$, we have

$$S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) = \int_{\mathbb{R}^3}\rho\log\rho - \left(\frac{N}{M}\log M - N\log\frac{N}{8\pi M}\right) + \frac{1}{M}\sum_{a=1}^{M}\int_{\mathbb{R}^3}|\mathbf{r}|\rho(\mathbf{r}+\mathbf{R}_a)\,d\mathbf{r},$$

which is a finite quantity for all $\rho \in X_+$. In the Hirschfeld-I, L-ISA, or NL-ISA settings, the sets $\mathcal{K}^0_{z_a}$ must contain functions decaying asymptotically not faster than $e^{-\alpha|\mathbf{r}|}$ (for some $\alpha > 0$) for (22) to be satisfied for all $\rho \in X_+$. The requirements on the sets $\mathcal{K}^0_{z_a}$ can be weakened by imposing more conditions on ρ. For instance, if we have $\int_{\mathbb{R}^3}|\mathbf{r}|^2\rho(\mathbf{r})\,d\mathbf{r} < \infty$, then condition (22) is satisfied as soon as the sets $\mathcal{K}^0_{z_a}$ contain functions decaying asymptotically not faster than a Gaussian.

Most of the iterative numerical schemes for solving (21) that have been proposed in the literature amounts to minimizing $S(\boldsymbol{\rho}|\boldsymbol{\rho}^0)$ alternatively with respect to each of the variables $\boldsymbol{\rho}^0$ and $\boldsymbol{\rho}$, under suitable constraints. They read as follows.

Generic iterative AIM algorithm:

- **Initialization**: Choose $\boldsymbol{\rho}^{0,(0)} \in \boldsymbol{\mathcal{K}}^0$ such that $s_{\mathrm{KL}}(\rho|\rho^{0,(0)}) < +\infty$.
- **Iteration** $m \geq 0$**:**

  Step 1: set

$$\boldsymbol{\rho}^{(m)} = \boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^{0,(m-1)}) = \underset{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}}{\arg\min}\, S(\boldsymbol{\rho}|\boldsymbol{\rho}^{0,(m-1)}), \tag{23}$$

  Step 2: find

$$\boldsymbol{\rho}^{0,(m)} \in \underset{\boldsymbol{\rho}^0\in\boldsymbol{\mathcal{K}}^0,\, \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)=\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^{(m)})}{\arg\min} S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^0). \tag{24}$$

Recall that the solution to (23) in Step 1 is explicit, since the expression of the function $\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0)$ is given by (16).

Due to the particular nature of $S(\cdot|\cdot)$, the solution to Step 2 can then be obtained by solving M independent and local problems of the form

$$\inf_{\rho_a^0\in\mathcal{K}^0_{z_a},\, \int_{\mathbb{R}^3}\rho_a^0=\int_{\mathbb{R}^3}\rho_a} s_{\mathrm{KL}}(\rho_a|\rho_a^0) \tag{25}$$

for $\rho_a = \rho_a^{(m-1)}$.

At the noticeable exception of the ISA method, solving the optimization problem in Step 2 for the various AIM methods presented above require evaluating the integrals

$$N_a^{(m)} := [\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^{(m)})]_a = \int_{\mathbb{R}^3} \rho_a^{(m)}(\mathbf{r})\, d\mathbf{r},$$

with $\boldsymbol{\rho}^{(m)} = \boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^{0,(m-1)})$. Step 2 therefore requires an efficient numerical quadrature scheme to evaluate integrals of the form

$$Q_{a,\rho,\mathbf{R}}(\boldsymbol{\rho}^0) := [\boldsymbol{\mathcal{N}}(\boldsymbol{G}_{\rho,\mathbf{R}}(\boldsymbol{\rho}^0))]_a = \int_{\mathbb{R}^3} \frac{\rho_a^0(\mathbf{r})}{\sum_{b=1}^N \rho_b^0(\mathbf{r}-\mathbf{R}_b+\mathbf{R}_a)}\rho(\mathbf{r}+\mathbf{R}_a)\, d\mathbf{r}, \tag{26}$$

(with the convention that the integrand is equal to zero when the denominator vanishes), for $\boldsymbol{\rho}^0 = (\rho_a^0)_{1\leq a\leq M}$ in the set $\boldsymbol{\mathcal{K}}^0$ of promolecule densities. The elements of the sets $\boldsymbol{\mathcal{K}}^0_{\mathrm{GISA}}$, $\boldsymbol{\mathcal{K}}^0_{\mathrm{Hirschfeld-I}}$ or $\boldsymbol{\mathcal{K}}^0_{\mathrm{MB-ISA}}$ of promolecule densities constructed from the GISA, Hirschfeld-I, and MB-ISA proatom densities respectively, are smooth, fast-decaying, radially symmetric functions. The integrals of the form (26) can therefore be efficiently computed using a one-dimensional quadrature scheme for the radial part and a Lebedev quadrature scheme for the angular part.

We successively present in details in the following sections how problems of the form (25) are solved for four different methods: the Hirschfeld-I method, the general ISA method, the finite-dimensional linear approximations of the ISA (L-ISA) method, the finite-dimensional nonlinear approximations of the ISA (NL-ISA) method.

Note that the abstract versions L-ISA and NL-ISA haven not been reported in the literature, although they are similar, but not identical, to the GISA and MB-ISA methods respectively. For L-ISA however, we can establish rigorous results (see Section 3) such as existence and uniqueness of the solution as well as global convergence of the AIM iteration scheme.

2.5.3 Hirschfeld-I iterations

For the Hirschfeld-I method [11], solving (25) is trivial since the set $\{\rho_a^0 \in \mathcal{K}^0_{z_a,\mathrm{Hirschfeld-I}} \mid \int_{\mathbb{R}^3} \rho_a^0 = N_a\}$ contains a single element, namely $\rho^0_{z_a,N_a}$. At the continuous level (i.e., without any discretization of the proatom densities), the Hirschfeld-I iterative procedure corresponding to the generic iterative AIM algorithm for

$$\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{Hirschfeld-I}} = \mathcal{K}^0_{z_1,\mathrm{Hirschfeld-I}} \times \cdots \times \mathcal{K}^0_{z_M,\mathrm{Hirschfeld-I}}$$

can therefore be formulated more explicitly as follows:

Hirschfeld-I algorithm:

- **Initialization:** choose an initial guess of charges $\mathbf{N}^{(0)} = (N_a^{(0)})_{1\le a\le M} \in (\mathbb{R}_+)^M$ so that $\sum_{a=1}^{N} N_a^{(0)} = N$.
- **Iteration** $m \ge 0$**:**

  set

$$\rho_a^{(m)}(\mathbf{r}) := \frac{\rho^0_{z_a,N_a^{(m)}}(\mathbf{r})}{\sum_{b=1}^{M} \rho^0_{z_b,N_b^{(m)}}(\mathbf{r}-\mathbf{R}_b+\mathbf{R}_a)} \rho(\mathbf{R}_a+\mathbf{r}), \tag{27}$$

  where

$$\rho^0_{z_a,N_a^{(m)}} = \left(\lceil N_a^{(m)}\rceil - N_a^{(m)}\right) \rho^0_{z_a,\lfloor N_a^{(m)}\rfloor} + \left(N_a^{(m)} - \lfloor N_a^{(m)}\rfloor\right) \rho^0_{z_a,\lceil N_a^{(m)}\rceil};$$

  compute

$$N_a^{(m+1)} := \int_{\mathbb{R}^3} \rho_a^{(m)}(\mathbf{r})\, d\mathbf{r}. \tag{28}$$

The atomic charges $N_a^{(m)}$ at iteration m are fractional (i.e. non-integer) in general. The global convergence property of this iterative scheme was observed numerically in [11] on a benchmark of 168 molecules. A convergence proof was proposed in [10].

From a numerical point of view, the Hirschfeld-I method can be seen as a fixed-point procedure of the form $\mathbf{N}^{(m+1)} = F_{\mathrm{Hirschfeld-I}}(\mathbf{N}^{(m)})$, where $F_{\mathrm{Hirschfeld-I}} : \mathbb{R}_+^M \to \mathbb{R}_+^M$ is the function defined by (27)-(28). Using $\mathbf{N} = (N_a)_{1\le a\le M} \in \mathbb{R}_+^M$ as main variables, it is not necessary to store in memory the functions $\rho_a^{(m)}$: they can be evaluated using (27) from ρ, $\mathbf{N}^{(m-1)}$, and the $\rho^0_{z,n}$'s. The implementation of the Hirschfeld-I method thus requires:

1. the pre-computation and storage of the radially symmetric functions $\rho^0_{z,n}$, for all chemical elements contained in the molecular system of interest, and integer values $n \le N_z^{\max}$. Note that these functions can be pre-computed once and for all with high accuracy and then be used for any molecular system;

2. repeated evaluations of integrals (26), where the ρ_a^0's are of the form $\rho^0_{z_a,N_a}$.

2.5.4 ISA iterations

The ISA method was originally introduced in [30, 31] as the iterative scheme described below, which was then interpreted in [23] as a specific instance of the generic iterative AIM algorithm for

$$\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}} := \left\{ \boldsymbol{\rho}^0 = (\rho^0_a)_{1\le a\le M} \in (X^r_+)^M \right\}.$$

For $\mathcal{K}^0_{z_a} = \mathcal{K}^0_{\mathrm{ISA}} = X^r_+$ and Step 2, the solution to problem (25), when it exists, has a particularly simple form. Indeed, denoting by

$$\begin{aligned}\langle \rho_a \rangle_s(\mathbf{r}) :&= \fint_{\mathbb{S}^2} \rho_a(|\mathbf{r}|\boldsymbol{\sigma})\, d\boldsymbol{\sigma} \\ &= \frac{1}{4\pi}\int_0^{\pi} \sin\theta \int_0^{2\pi} \rho_a(r\sin\theta\cos\phi\,\mathbf{e}_x + r\sin\theta\sin\phi\,\mathbf{e}_y + r\cos\theta\,\mathbf{e}_z)\, d\phi\, d\theta,\end{aligned} \tag{29}$$

with $\fint_{\mathbb{S}^2} := \frac{1}{4\pi}\int_{\mathbb{S}^2}$, the spherical average of ρ_a, we have the following result.

Lemma 5. *Let $\rho_a \in X_+$. Then, the constrained optimization problem*

$$\inf_{\rho^0_a \in \mathcal{K}^0_{\mathrm{ISA}},\ \int_{\mathbb{R}^3}\rho^0_a = \int_{\mathbb{R}^3}\rho_a} s_{\mathrm{KL}}(\rho_a|\rho^0_a) \tag{30}$$

is well-posed and its solution is $\langle \rho_a \rangle_s$.

It is convenient to formulate the ISA method in terms of the radial functions

$$w_a(r) := \fint_{\mathbb{S}^2} \rho_a(r\boldsymbol{\sigma})\, d\boldsymbol{\sigma}, \tag{31}$$

which belong to the closed convex cone

$$Y_+ := \{w \in L^1(\mathbb{R}_+; (r^2+r^3)\,dr) \cap L^\infty(\mathbb{R}_+) \mid w \ge 0 \text{ a.e.}, \lim_{r\to+\infty} w(r) = 0\}$$

of the Banach space $Y := L^1(\mathbb{R}_+; (r^2+r^3)\,dr) \cap L^\infty(\mathbb{R}_+)$, and are such that

$$\forall r > 0,\ \forall \boldsymbol{\sigma} \in \mathbb{S}^2, \quad \langle \rho_a \rangle_s(r\boldsymbol{\sigma}) = w_a(r).$$

The functions w_a are the main variables used in the practical implementation of the ISA method. At the continuous level, the ISA algorithm can thus be formulated as follows:

ISA algorithm:

- **Initialization:** choose an initial guess for $\mathbf{w}^{(0)} := (w^{(0)}_a)_{1\le a\le M} \in Y^M_+$ such that $\sum_{a=1}^M w_a(|\mathbf{r}-\mathbf{R}_a|) > 0$ in $\mathbb{R}^3$.
- **Iteration** $m \ge 1$**:**

  Step 1: set

  $$\rho^{(m)}_a(\mathbf{r}) := \frac{w^{(m-1)}_a(|\mathbf{r}|)}{\sum_{b=1}^M w^{(m-1)}_b(|\mathbf{r}-\mathbf{R}_b+\mathbf{R}_a|)}\rho(\mathbf{R}_a+\mathbf{r}),$$

  Step 2: compute

  $$w^{(m)}_a(r) := \fint_{\mathbb{S}^2} \rho^{(m)}_a(r\boldsymbol{\sigma})\, d\boldsymbol{\sigma}. \tag{32}$$

Introducing the linear map $X^M \ni \boldsymbol{\rho} \mapsto \langle \boldsymbol{\rho} \rangle_s \in X^M$ where for all $\boldsymbol{\rho} = (\rho_a)_{1 \le a \le M} \in X^M$, $\langle \boldsymbol{\rho} \rangle_s := (\langle \rho_a \rangle_s)_{1 \le a \le M}$ with $\langle \rho_a \rangle_s$ defined by (29), the ISA optimization problem can be written as

$$\min_{\boldsymbol{\rho} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho} | \langle \boldsymbol{\rho} \rangle_s) \qquad \text{(ISA optimization problem).} \tag{33}$$

A practical implementation of ISA can then be obtained by

1. discretizing the radially symmetric functions $w_a^{(m-1)}$ on a suitable one-dimensional grid;
2. computing the spherical averages (32) for each a and each grid point of the radial grid, using e.g. Lebedev integration method.

More details about the practical aspects of the implementation will be given in Section 5.

2.5.5 Linear approximation of the ISA (L-ISA) method

Instead of working with the three-dimensional, but radially symmetric, functions given in $\mathcal{K}_z^0 = \mathcal{K}_{z,\mathrm{L-ISA}}^0$, it is more convenient to work with their generating function defined on $\mathbb{R}_+$. For this purpose, let us denote by $\widetilde{g}_{z_a,k} : \mathbb{R}_+ \to \mathbb{R}_+$ the function such that $g_{z_a,k}(\mathbf{r}) = \widetilde{g}_{z_a,k}(|\mathbf{r}|)$ for all $\mathbf{r} \in \mathbb{R}^3$, $1 \le a \le M$ and $1 \le k \le m_{z_a}$. Let us also denote by

$$V := \mathrm{Span}\left\{\widetilde{g}_{z_a,k},\ 1 \le a \le M,\ 1 \le k \le m_{z_a}\right\}, \quad V_+ := \mathrm{Span}_+\left\{\widetilde{g}_{z_a,k},\ 1 \le a \le M,\ 1 \le k \le m_{z_a}\right\}$$

(where the notation Span_+ denotes the set of linear combinations with non-negative coefficients of the $\widetilde{g}_{z_a,k}$'s). Then, it holds that $V \subset Y$, $V_+ \subset Y_+$ and the L-ISA algorithm can be rewritten as follows:

L-ISA algorithm:

- **Initialization:** choose an initial guess for $\mathbf{v}^{(0)} := (v_a^{(0)})_{1 \le a \le M} \in V_+^M$ such that $\sum_{b=1}^{M} v_b(|\mathbf{r} - \mathbf{R}_b|) > 0$ in $\mathbb{R}^3$.
- **Iteration** $m \ge 1$**:**

  Step 1: set

$$\rho_a^{(m)}(\mathbf{r}) := \frac{v_a^{(m-1)}(|\mathbf{r}|)}{\sum_{b=1}^{M} v_b^{(m-1)}(|\mathbf{r} - \mathbf{R}_b + \mathbf{R}_a|)} \rho(\mathbf{R}_a + \mathbf{r}), \tag{34}$$

  Step 2: compute

$$v_a^{(m)} \in \underset{\substack{v_a \in V_+, \\ 4\pi \int_{\mathbb{R}_+} v_a(r) r^2\, dr = N_a^{(m)}}}{\operatorname{argmin}} \int_{\mathbb{R}^3} \rho_a^{(m)}(\mathbf{r}) \log\left(\frac{\rho_a^{(m)}(\mathbf{r})}{v_a(|\boldsymbol{r}|)}\right) d\mathbf{r}, \tag{35}$$

  with

$$N_a^{(m)} := \int_{\mathbb{R}^3} \rho_a^{(m)}(\boldsymbol{r})\, d\boldsymbol{r}.$$

In this case, the subproblems (25) (and equivalently subproblems (35) with $\rho_a = \rho_a^{(m)}$) to be solved in Step 2 are of the form

$$\inf_{\substack{v_a \in V_+, \\ 4\pi \int_{\mathbb{R}_+} v_a(r) r^2 \, dr = N_a}} \int_{\mathbb{R}^3} \rho_a(\mathbf{r}) \log \left(\frac{\rho_a(\mathbf{r})}{v_a(|\boldsymbol{r}|)} \right) d\mathbf{r}, \tag{36}$$

with $N_a := \int_{\mathbb{R}^3} \rho_a$. The existence of minimizers to this problem is discussed in Lemma 11. Problem (36) can be rewritten equivalently, since the functions $g_{z_a,k}$ are L^1-normalized, as

$$\inf_{\substack{\mathbf{c}_a = (c_{a,1}, \cdots, c_{a,m_{z_a}}) \in \mathbb{R}_+^{m_{z_a}} \\ \sum_{k=1}^{m_{z_a}} c_{a,k} = N_a}} \int_{\mathbb{R}^3} \rho_a(\mathbf{r}) \log \left(\frac{\rho_a(\mathbf{r})}{\sum_{k=1}^{m_{z_a}} c_{a,k} \, g_{z_a,k}(\mathbf{r})} \right) d\mathbf{r}. \tag{37}$$

Note that the minimization problem (37) differs from the original GISA method [57] where an L^2 norm distance between $v_a = \sum_{k=1}^{m_{z_a}} c_{a,k} \, g_{z_a,k}$ and the spherical average of ρ_a (namely $\rho_a^{(m)}$, at iteration m) is minimized, though under the same constraint $\sum_{k=1}^{m_{z_a}} c_{a,k} = N_a$ (see Section 2.6).

Using the function w_a defined in (31), we see that

$$\int_{\mathbb{R}^3} \rho_a(\mathbf{r}) \log \left(\frac{\rho_a(\mathbf{r})}{\sum_{k=1}^{m_{z_a}} c_{a,k} \, g_{z_a,k}(\mathbf{r})} \right) d\mathbf{r} = \int_{\mathbb{R}^3} \rho_a(\mathbf{r}) \log \left(\rho_a(\mathbf{r}) \right) d\mathbf{r} - 4\pi \int_0^{+\infty} r^2 w_a(r) \log \left(\sum_{k=1}^{m_{z_a}} c_{a,k} \, \widetilde{g}_{z_a,k}(r) \right) dr.$$

It follows that the minimizers of (37) can be obtained by solving the problem

$$\inf_{\substack{\mathbf{c}_a = (c_{a,1}, \cdots, c_{a,m_{z_a}}) \in \mathbb{R}_+^{m_{z_a}} \\ \sum_{k=1}^{m_{z_a}} c_{a,k} = N_a}} F_a(\mathbf{c}_a), \tag{38}$$

where

$$F_a(\mathbf{c}_a) := - \int_0^{+\infty} r^2 w_a(r) \log \left(\sum_{k=1}^{m_{z_a}} c_{a,k} \, \widetilde{g}_{z_a,k}(r) \right) dr. \tag{39}$$

For functions $g_{z_a,k}$ decaying extremely fast at infinity, it may happen that $F_a(\mathbf{c}_a) = +\infty$ for w_a's corresponding to physically admissible ρ_a's. On the other hand, if the $\widetilde{g}_{z_a,k}$ decay at most exponentially fast (resp. not faster than a Gaussian function) and the first-order moments (resp. the second-order moments) of the function ρ to be decomposed is finite, then the function F_a is

strictly convex and continuous on the simplex $\left\{ \mathbf{c}_a = (c_{a,1}, \cdots, c_{a,m_{z_a}}) \in \mathbb{R}_+^{m_{z_a}} \;\middle|\; \sum_{k=1}^{m_{z_a}} c_{a,k} = N_a \right\}$ for all $N_a > 0$, and therefore has a unique minimizer. In addition, it is smooth, so that the minimizer can be computed efficiently for small values of m_{z_a} by standard numerical optimization algorithms. The components of the gradient and Hessian of F_a are given respectively for all $1 \leq i, j \leq m_{z_a}$ by

$$\frac{\partial F_a}{\partial c_{a,i}}(\mathbf{c}_a) = -\int_0^{+\infty} r^2 w_a(r) \frac{\widetilde{g}_{z_a,i}(r)}{\sum_{k=1}^{m_{z_a}} c_{a,k} \widetilde{g}_{z_a,k}(r)} \, dr,$$

$$\frac{\partial^2 F_a}{\partial c_{a,i} \partial c_{a,j}}(\mathbf{c}_a) = \int_0^{+\infty} r^2 w_a(r) \frac{\widetilde{g}_{z_a,i}(r) \widetilde{g}_{z_a,j}(r)}{\left(\sum_{k=1}^{m_{z_a}} c_{a,k} \, \widetilde{g}_{z_a,k}(r) \right)^2} \, dr.$$

2.5.6 Nonlinear approximation of the ISA (NL-ISA)

Proceeding as in the previous section, the subproblems (25) to be solved in Step 2 are now of the form

$$\inf_{\substack{\mathbf{c}_a \in \mathbb{R}_+^{m_{z_a}}, \boldsymbol{\alpha}_a \in (0,+\infty)^{m_{z_a}} \\ \sum_{k=1}^{m_{z_a}} c_{a,k} = N_a}} F_a(\mathbf{c}_a, \boldsymbol{\alpha}_a), \tag{40}$$

denoting $\mathbf{c}_a = (c_{a,1}, \cdots, c_{a,m_{z_a}})$, $\boldsymbol{\alpha}_a = (\alpha_{a,1}, \cdots, \alpha_{a,m_{z_a}})$ and where

$$F_a(\mathbf{c}_a, \boldsymbol{\alpha}_a) := -\int_0^{+\infty} r^2 w_a(r) \log \left(\sum_{k=1}^{m_{z_a}} c_{a,k} \, \widetilde{g}_{z_a,k,\alpha_{a,k}}(r) \right) dr.$$

Still for the sake of simplicity, we assume that the $\alpha_{a,k}$'s are positive real numbers, but more complicated settings can be considered as well.

Not much can be said about the optimization (40) without additional assumptions on the functions $\widetilde{g}_{z_a,k,\alpha_{a,k}}$. It may have no solution for some admissible w_a's, and multiple solutions for other admissible w_a's.

2.6 GISA iterations

The aim of this section is to describe in details the original GISA method. The GISA iterations can be formalized as follows.

GISA algorithm:

- **Initialization:** Choose $\boldsymbol{\rho}^{0,(0)} \in \boldsymbol{\mathcal{K}}^0_{\text{GISA}}$.
- **Iteration** $m \geq 1$**:**

  Step 1: Find $\boldsymbol{\rho}^{(m)} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ solution to

$$\boldsymbol{\rho}^{(m)} = \underset{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}}{\arg\min}\, S(\boldsymbol{\rho}|\boldsymbol{\rho}^{0,(m-1)}). \tag{41}$$

  Step 2: Find $\boldsymbol{\rho}^{0,(m)} \in \boldsymbol{\mathcal{K}}^0_{\text{GISA}}$ solution to

$$\boldsymbol{\rho}^{0,(m)} = \underset{\boldsymbol{\rho}^0\in\boldsymbol{\mathcal{K}}^0_{\text{GISA}}\,|\,\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)=\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^{(m)})}{\arg\min} \|\boldsymbol{\rho}^0 - \boldsymbol{\rho}^{(m)}\|^2_{L^2}. \tag{42}$$

The solution to (41) is explicit, while (42) can be split into the M independent minimization problems

$$\rho_a^{0,(m)} = \underset{\rho_a^0\in\mathcal{K}^0_{z_a,\text{GISA}}\,|\,\int_{\mathbb{R}^3}\rho_a^0=\int_{\mathbb{R}^3}\rho_a^{(m)}}{\arg\min} \|\rho_a^{(0)} - \rho_a^{(m)}\|^2_{L^2}. \tag{43}$$

It is convenient to reformulate the GISA iterations using as main variables the coefficients $\mathbf{c} = (\mathbf{c}_a)_{1\leq a\leq M} \in \mathbb{R}^p_+$ with $\mathbf{c}_a = (c_{a,k})_{1\leq k\leq m_{z_a}} \in \mathbb{R}^{m_{z_a}}_+$ and $p = \sum_{a=1}^M m_{z_a}$ (see Eq. (8)). The iterations (41)-(42) can be rewritten as

$$\rho_a^{(m)}(\mathbf{r}) = \frac{\displaystyle\sum_{k=1}^{m_{z_a}} c_{a,k}^{(m-1)}\,\zeta_{\alpha_{a,k}}(\mathbf{r})}{\displaystyle\sum_{b=1}^{M}\sum_{k=1}^{m_{z_b}} c_{b,k}^{(m-1)}\,\zeta_{\alpha_{b,k}}(\mathbf{r}+\mathbf{R}_a-\mathbf{R}_b)}\rho(\mathbf{r}+\mathbf{R}_a), \tag{44}$$

and

$$\mathbf{c}_a^{(m)} = \underset{\boldsymbol{c}_a\in\mathbb{R}^{m_{z_a}}_+,\ \sum_{k=1}^{m_{z_a}} c_{a,k} = \int_{\mathbb{R}^3}\rho_a^{(m)}}{\arg\min} \left\| \sum_{k=1}^{m_{z_a}} c_{a,k}\,\zeta_{\alpha_{a,k}} - \rho_a^{(m)} \right\|^2_{L^2} \tag{45}$$

$$= \underset{\boldsymbol{c}_a\in\mathbb{R}^{m_{z_a}}_+,\ \sum_{k=1}^{m_{z_a}} c_{a,k} = \int_{\mathbb{R}^3}\rho_a^{(m)}}{\arg\min} \left(\frac{1}{2}\boldsymbol{c}_a^T\boldsymbol{S}_{z_a}\boldsymbol{c}_a - \boldsymbol{c}_a^T\boldsymbol{b}_a^{(m)}\right), \tag{46}$$

with

$$[S_{z_a}]_{k\ell} = 2\int_{\mathbb{R}^3}\zeta_{\alpha_{a,k}}(\mathbf{r})\,\zeta_{\alpha_{a,\ell}}(\mathbf{r})\,d\mathbf{r} = \frac{2}{\pi\sqrt{\pi}}\frac{(\alpha_{a,k}\alpha_{a,\ell})^{\frac{3}{2}}}{\left(\sqrt{\alpha_{a,k}+\alpha_{a,\ell}}\right)^3},$$

$$b_{a,k}^{(m)} = \int_{\mathbb{R}^3}\zeta_{\alpha_{a,k}}(\mathbf{r})\,\rho_a^{(m)}(\mathbf{r})\,d\mathbf{r}.$$

Again, the functions $\rho_a^{(m)}$ do not need to be stored in memory. The integrals $\int_{\mathbb{R}^3} \rho_a^{(m)}$ and the entries of the vectors $\boldsymbol{b}_a^{(m)}$ can be computed using quadrature formulas from the vector $\mathbf{c}_a^{(m-1)}$ and the function ρ. In practice, m_{z_a}'s are small so that (45) can be solved very easily once these integrals have been computed, using standard routines for quadratic programming problems [20, 21].

The existence of a fixed point to the GISA iterations can be proved using Brouwer fixed-point theorem. On the other hand, the convergence of the GISA iterations for any initial guess is not guaranteed.

2.7 MB-ISA iterations

The MB-ISA method [59] was originally defined through a Lagrangian that does not seem to be canonically associated to a well-defined constrained optimization problem. It is more satisfactory from a mathematical point of view, to reformulate it as follows:

1. the density ρ is decomposed as a sum of atomic-shell densities. The set of admissible atomic-shell-in-molecule (ASIM) decompositions of the density ρ is

$$\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R},\boldsymbol{z}} = \left\{ \boldsymbol{\rho} = (\rho_{a,k})_{1\leq a\leq M,\, 1\leq k\leq m_{z_a}} \in X_+^M \;\middle|\; \sum_{a=1}^{M}\sum_{k=1}^{m_{z_a}} \rho_{a,k}(\cdot - \mathbf{R}_a) = \rho \right\}.$$

   The integer m_{z_a} is the number of electronic shells in atom a; it is a function of the atomic number z_a (hence the subscript $\boldsymbol{z} = (z_1, \ldots, z_M) \in (\mathbb{N}^*)^M$ collecting the atomic numbers of the atoms contained in the molecular system of interest);

2. the set of proatomic-shell densities is defined as

$$\mathcal{K}^0_{\mathrm{MB-ISA}} = \left\{ \rho^0(\mathbf{r}) = c\frac{\alpha^3}{8\pi} e^{-\alpha|\mathbf{r}|},\ c \geq 0,\ \alpha > 0 \right\},$$

   (note that this set is independent of the chemical element and the shell) and the set of admissible pro-ASIM densities as

$$\boldsymbol{\mathcal{K}}^0_{\boldsymbol{z}} = \left\{ \boldsymbol{\rho}^0 = (\rho^0_{a,k})_{1\leq a\leq M,\, 1\leq k\leq m_{z_a}},\ \rho^0_{a,k} \in \mathcal{K}^0_{\mathrm{MB-ISA}} \right\};$$

3. the atomic-shell relative entropy is defined as

$$S_{\mathrm{sh}}(\boldsymbol{\rho}|\boldsymbol{\rho}^0) := \sum_{a=1}^{M}\sum_{k=1}^{m_{z_a}} s_{\mathrm{KL}}(\rho_{a,k}|\rho^0_{a,k});$$

4. the optimal MB-ISA decomposition is finally obtained by solving the constrained minimization problem

$$\inf_{(\boldsymbol{\rho},\boldsymbol{\rho}_0)\in\boldsymbol{\mathcal{C}}_{\rho,\mathbf{R},\boldsymbol{z}}} S_{\mathrm{sh}}(\boldsymbol{\rho}|\boldsymbol{\rho}^0), \tag{47}$$

   where

$$\boldsymbol{\mathcal{C}}_{\rho,\mathbf{R},\boldsymbol{z}} = \left\{ (\boldsymbol{\rho},\boldsymbol{\rho}^0) \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R},\boldsymbol{z}} \times \boldsymbol{\mathcal{K}}^0_{\boldsymbol{z}} \mid \boldsymbol{\mathcal{M}}_0(\boldsymbol{\rho}) = \boldsymbol{\mathcal{M}}_0(\boldsymbol{\rho}^0),\ \boldsymbol{\mathcal{M}}_1(\boldsymbol{\rho}) = \boldsymbol{\mathcal{M}}_1(\boldsymbol{\rho}^0) \right\}.$$

   Denoting by $K = \sum_{a=1}^{M} m_{z_a}$ and

$$X_n := \left\{ \rho \in X \;\middle|\; \int_{\mathbb{R}^3} |\mathbf{r}|^n |\rho(\mathbf{r})|\, d\mathbf{r} < \infty \right\},$$

the linear maps $\boldsymbol{\mathcal{M}}_n : X_n^K \to \mathbb{R}^K$ are defined by

$$\forall \boldsymbol{\rho} = (\rho_{a,k})_{1\le a\le M,\, 1\le k\le m_{z_a}} \in X_n^K,\ [\boldsymbol{\mathcal{M}}_n(\boldsymbol{\rho})]_{a,k} = \int_{\mathbb{R}^3} |\mathbf{r}|^n \rho_{a,k}(\mathbf{r})\, d\mathbf{r}.$$

Note that $X_0 = X$ and $\boldsymbol{\mathcal{M}}_0 = \boldsymbol{\mathcal{N}}$. As for any $c \in \mathbb{R}$,

$$\int_{\mathbb{R}^3} c\frac{\alpha^3}{8\pi} e^{-\alpha|\mathbf{r}|}\, d\mathbf{r} = c \quad \text{and} \quad \int_{\mathbb{R}^3} |\mathbf{r}| c\frac{\alpha^3}{8\pi} e^{-\alpha|\mathbf{r}|}\, d\mathbf{r} = \frac{3c}{\alpha},$$

the function

$$\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\boldsymbol{z}}}(\boldsymbol{\rho}) := \operatorname{argmin}\big\{S_{\mathrm{sh}}(\boldsymbol{\rho}|\boldsymbol{\rho}^0),\ \boldsymbol{\rho}^0 \in \boldsymbol{\mathcal{K}}^0_{\boldsymbol{z}},\ \boldsymbol{\mathcal{M}}_0(\boldsymbol{\rho}^0) = \boldsymbol{\mathcal{M}}_0(\boldsymbol{\rho}),\ \boldsymbol{\mathcal{M}}_1(\boldsymbol{\rho}^0) = \boldsymbol{\mathcal{M}}_1(\boldsymbol{\rho})\big\},$$

has an explicit expression: for all $\boldsymbol{\rho} = (\rho_{a,k})_{1\le a\le M,\, 1\le k\le m_{z_a}} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R},\boldsymbol{z}}$,

$$[\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\boldsymbol{z}}}(\boldsymbol{\rho})]_{a,k}(\mathbf{r}) = c(\rho_{a,k})\frac{\alpha(\rho_{a,k})^3}{8\pi} e^{-\alpha(\rho_{a,k})|\mathbf{r}|},$$

with

$$c(\rho_{a,k}) = \int_{\mathbb{R}^3} \rho_{a,k}(\mathbf{r})\, dr \quad \text{and} \quad \alpha(\rho_{a,k}) = 3\, c(\rho_{a,k}) \left(\int_{\mathbb{R}^3} |\mathbf{r}|\rho_{a,k}(\mathbf{r})\, dr\right)^{-1}.$$

MB-ISA can be reformulated in an easily implementable iterative algorithm in which the main variable is $\mathbf{y} := (c_{a,k}, \alpha_{a,k})_{1\le a\le M,\, 1\le k\le m_{z_a}} \in \mathbb{R}^K_+$:

MB-ISA algorithm:

- **Initialization:** choose $(c^{(0)}_{a,k})_{1\le a\le M, 1\le k\le m_{z_a}}$ and $(\alpha^{(0)}_{a,k})_{1\le a\le M, 1\le k\le m_{z_a}}$ – see section 5.1 for details.
- **Iteration** $m \ge 1$**:** set

$$c^{(m)}_{a,k} = \int_{\mathbb{R}^3} \frac{c^{(m-1)}_{a,k}\left(\alpha^{(m-1)}_{a,k}\right)^3 e^{-\alpha^{(m-1)}_{a,k}|\mathbf{r}|}}{\sum_{b=1}^{M}\left(\sum_{\ell=1}^{K_b} c^{(m-1)}_{b,\ell}\left(\alpha^{(m-1)}_{b,\ell}\right)^3 e^{-\alpha^{(m-1)}_{b,\ell}|\mathbf{r}+\mathbf{R}_a-\mathbf{R_b}|}\right)}\rho(\mathbf{R}_a+\mathbf{r})d\mathbf{r}, \tag{48}$$

$$\alpha^{(m)}_{a,k} = 3\, c^{(m)}_{a,k}\left(\int_{\mathbb{R}^3} \frac{c^{(m-1)}_{a,k}\left(\alpha^{(m-1)}_{a,k}\right)^3 e^{-\alpha^{(m-1)}_{a,k}|\mathbf{r}|}}{\sum_{b=1}^{M}\left(\sum_{\ell=1}^{K_b} c^{(m-1)}_{b,\ell}\left(\alpha^{(m-1)}_{b,\ell}\right)^3 e^{-\alpha^{(m-1)}_{b,\ell}|\mathbf{r}+\mathbf{R}_a-\mathbf{R_b}|}\right)}|\mathbf{r}|\rho(\mathbf{R}_a+\mathbf{r})d\mathbf{r}\right)^{-1}, \tag{49}$$

which only requires repeated evaluations of integrals of the form (26) (with $\rho(\mathbf{R}_a+\mathbf{r})$ replaced by $|\mathbf{r}|\rho(\mathbf{R}_a+\mathbf{r})$ in the evaluation of (49)).

3 Mathematical analysis of AIM decomposition methods

In this section, we make a specific focus on the AIM decomposition methods falling into the scope of the unified formulation (21): find $\left(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}}\right) \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0$ solution to

$$\inf_{\substack{\left(\boldsymbol{\rho},\boldsymbol{\rho}^0\right) \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0 \\ \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)}} S\left(\boldsymbol{\rho}|\boldsymbol{\rho}^0\right), \tag{50}$$

where $\boldsymbol{\mathcal{K}}^0$ is a non-empty subset of $(X_+^r)^M$. Note that the only freedom to change the method is the set $\boldsymbol{\mathcal{K}}^0$ and the remaining mathematical structure remains identical.

For the ISA and L-ISA method, the set $\boldsymbol{\mathcal{K}}^0$ is convex and closed for the topology of X^M. Consequently, (50) is a strictly convex problem, so that its solution, if it exists, is unique. On the other hand, the existence of a minimizer is not obvious *a priori* due to possible loss of compactness; this issue is addressed in Theorem 6.

For the Hischfeld-I method, the existence of a minimizer follows from the very simple structure of $\boldsymbol{\mathcal{K}}^0_{\text{Hirschfeld}-\text{I}}$ by a simple compactness argument, while the uniqueness of the AIM decomposition is not guaranteed due to the non-convexity of $\boldsymbol{\mathcal{K}}^0_{\text{Hirschfeld}-\text{I}}$. For NL-ISA, neither existence, nor uniqueness, can be proved in the absence of further assumptions on ρ and the functions $\widetilde{g}_{z_a,k,\alpha_{a,k}}$.

In the rest of the section, we focus on ISA and L-ISA. Let us emphasize that although the sets $\boldsymbol{\mathcal{K}}^0_{\text{L}-\text{ISA}}$ and $\boldsymbol{\mathcal{K}}^0_{\text{GISA}}$ have the same mathematical structure, the arguments below cannot be applied to GISA since the latter approach does not fit into the framework (50). The existence and uniqueness of a minimizer to the corresponding optimization problem is established in Theorem 6. The aim of Theorem 7 is to study the convergence properties of ISA and L-ISA. We prove that the objective functional is non-increasing along the iterations and that the L^2-distance between to successive iterates tends to zero as the iteration number tends to infinity. In the L-ISA case, we prove the convergence of the L-ISA algorithm towards the unique minimizer of the L-ISA optimization problem. Some additional mathematical properties of the minimizer of the ISA optimization problem in the diatomic case are collected in Propositions 8 and 9. All the proofs are postponed until Section 6.

Theorem 6. *Let $\rho \in X_+$. Then,*

1. *For $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\text{ISA}}$, problem (50) admits a unique minimizer $\left(\boldsymbol{\rho}^{\text{opt}}, \boldsymbol{\rho}^{0,\text{opt}}\right) \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0_{\text{ISA}}$.*
2. *The same holds for $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\text{L}-\text{ISA}}$ under the following additional assumption*

$$\forall 1 \leq a \leq M, \quad \rho(\,\cdot + \mathbf{R}_a) \log\left(\min_{1\leq k\leq m_{z_a}} g_{z_a,k}(\cdot)\right) \in L^1(\mathbb{R}^3). \tag{51}$$

3. *In both cases (ISA and L-ISA), $\boldsymbol{\rho}^{\text{opt}} := \left(\rho_a^{\text{opt}}\right)_{1\leq a\leq M}$ and $\boldsymbol{\rho}^{0,\text{opt}} := \left(\rho_a^{0,\text{opt}}\right)_{1\leq a\leq M}$ satisfy the Hirschfeld relation*

$$\forall 1 \leq a \leq M, \quad \rho_a^{\text{opt}}(\boldsymbol{r}) = \frac{\rho_a^{0,\text{opt}}(\boldsymbol{r})}{\displaystyle\sum_{b=1}^{M} \rho_b^{0,\text{opt}}(\boldsymbol{r} - \mathbf{R}_b + \mathbf{R}_a)}\, \rho(\boldsymbol{r} + \mathbf{R}_a), \tag{52}$$

   with the convention that $\rho_a^{\text{opt}}(\boldsymbol{r}) = 0$ whenever $\rho_a^{0,\text{opt}}(\boldsymbol{r}) = 0$.
4. *In the ISA case, we have in addition*

$$\rho_a^{0,\text{opt}} = \langle \rho_a^{\text{opt}} \rangle_{\text{s}}. \tag{53}$$

At first sight, it is not clear at all that the fixed-point iterations (23)-(24) for solving problem (50) should converge. Indeed, the constraint $\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)$ is not taken into account in

the minimization subproblem (24), and it is therefore not clear *a priori* that the algorithm (23)-(24) possesses a Lyapunov functional (i.e. a functional which decreases at each iteration, a very useful tool in convergence proofs). It turns out that the function S itself is in fact a Lyapunov functional: $S\left(\boldsymbol{\rho}^{(m+1)}|\boldsymbol{\rho}^{0,(m+1)}\right) \leq S\left(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)}\right)$ for all m. This is a consequence of the special properties of the Kullback-Leibler divergence, which gives rise to the following result.

Theorem 7. *Let $\rho \in X_+$ be such that $\rho > 0$ almost everywhere, and $\boldsymbol{\rho}^{0,(0)} = (\rho_a^{0,(0)})_{1\leq a\leq M} \in \boldsymbol{\mathcal{K}}^0$ (with $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$ or $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}$) be an initial guess such that $\rho_a^{0,(0)} > 0$ almost everywhere for all $1 \leq a \leq M$, and $s_{\mathrm{KL}}(\rho|\rho^{0,(0)}) < +\infty$, where $\rho^{0,(m)} = \sum_{a=1}^M \rho_a^{0,(m)}(\cdot - \mathbf{R}_a)$. In the L-ISA case, we assume in addition that (51) is satisfied and all the $g_{z_a,k}$'s are positive almost everywhere. Then*

1. *For both ISA and L-ISA, the AIM iterations (23)-(24) are well-defined and S is a Lyapunov functional of the algorithm. More precisely, we have for all $m \geq 1$, $(\boldsymbol{\rho}^{(m)}, \boldsymbol{\rho}^{0,(m)}) \in \boldsymbol{\mathcal{C}}_{\rho,\mathbf{R}}$ and the estimates*

$$S(\boldsymbol{\rho}^{(m-1)}|\boldsymbol{\rho}^{0,(m-1)}) - S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)}) \geq \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^M \left\|\rho_a^{(m)} - \rho_a^{(m-1)}\right\|_{L^2}^2 \geq 0, \tag{54}$$

and

$$s_{\mathrm{KL}}(\rho|\rho^{0,(m-1)}) - s_{\mathrm{KL}}(\rho|\rho^{0,(m)}) \geq \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^M \left\|\rho_a^{(m+1)} - \rho_a^{(m)}\right\|_{L^2}^2 \geq 0. \tag{55}$$

As a consequence, it holds that

$$\boldsymbol{\rho}^{(m)} - \boldsymbol{\rho}^{(m-1)} \underset{m\to+\infty}{\longrightarrow} 0 \quad \text{strongly in } L^2(\mathbb{R}^3)^M. \tag{56}$$

2. *For L-ISA, the sequence $(\boldsymbol{\rho}^{0,(m)}, \boldsymbol{\rho}^{(m)})_{m\in\mathbb{N}}$ converges to the unique minimizer $(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}})$ of (50) in some sense. More precisely, $(\boldsymbol{\rho}^{0,(m)})_{m\in\mathbb{N}}$ converges strongly to $\boldsymbol{\rho}^{0,\mathrm{opt}}$ in $L^p(\mathbb{R}^3)$ for any $1 \leq p \leq +\infty$, and $(\boldsymbol{\rho}^{(m)})_{m\in\mathbb{N}}$ converges to $\boldsymbol{\rho}^{\mathrm{opt}}$ for the weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, for the weak topology of $L^p(\mathbb{R}^3)^M$ for any $1 < p < \infty$, and for the strong topology of $L^\infty_{\mathrm{loc}}(\mathbb{R}^3)$.*

Here, $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)$ denotes the Banach space of the bounded (signed) Radon measures on $\mathbb{R}^3$ endowed with the total variation norm.

We finally establish additional mathematical properties of the ISA AIM decomposition $\boldsymbol{\rho}^{\mathrm{opt}} = (\rho_1^{\mathrm{opt}}, \rho_2^{\mathrm{opt}})$ of diatomic systems ($\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$ and $M = 2$). We first show that one of the atomic densities ρ_a^{opt} is identically equal to zero if and only the total density ρ is a radial function centered on the other atom.

Proposition 8. *Let us assume that $M = 2$ and $\rho \in X_+$ is continuous and positive on $\mathbb{R}^3$. Let $\left(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}}\right)$ be the unique minimizer of problem (33) for $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$. Then, $\rho_1^{\mathrm{opt}} = 0$ if and only if $\rho(\cdot + \mathbf{R}_2) = \rho_2^{\mathrm{opt}} = \langle\rho_2^{\mathrm{opt}}\rangle_s$.*

We also have the following result, which proves stronger regularity results on $\boldsymbol{\rho}$ under appropriate assumptions on the regularity of ρ.

Proposition 9. *Assume that $M = 2$ and $\rho \in X_+ \cap C^{0,1}(\mathbb{R}^3)$. Let $(\boldsymbol{\rho}^{\mathrm{opt}}, \boldsymbol{\rho}^{0,\mathrm{opt}})$ be the unique minimizer of problem (50) for $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$. For $a = 1, 2$, let $w_a \in Y_+$ be the function defined by*

$$\forall r > 0, \quad \forall \boldsymbol{\sigma} \in \mathbb{S}^2, \quad w_a(r) := \langle\rho_a^{\mathrm{opt}}\rangle_s(r\boldsymbol{\sigma}).$$

Then,

1. *for almost all* $r \geq 0$, $w_a(r) \in \{0, w_a^+(r)\}$, *where the functions* $w_a^+ : \mathbb{R}_+ \to \mathbb{R}_+$ *are bounded by* $M_\rho := \max_{\mathbb{R}^3} \rho$, *continuous on* $(0, +\infty)$, *and Lipschitz in the neighborhood of any point of* $(0, +\infty)$ *at which they are positive;*

2. *if in addition, the density* ρ *is* C^k *away from the centers* $\mathbf{R}_a$ *for* $a = 1, 2$ *and the functions* w_a *are essentially bounded away from zero on any compact subset of* $[0, +\infty)$, *then the functions* w_a *are Lipschitz on* $[0, +\infty)$ *and* C^k *on* $(0, R) \cup (R, +\infty)$ *with* $R := |\mathbf{R}_a - \mathbf{R}_b|$.

4 Distributed Multipole Analysis

Other schemes allow to represent the electronic density as a sum of atom-centered contributions. Mulliken population analysis [38], and Distributed Multipole Analysis (DMA) [52, 54] – which can be viewed as a generalization of Mulliken scheme to higher-order multipole moments – belong to the latter. These schemes are fully basis-set oriented, in the sense that they use as a starting point a decomposition of the density of the form

$$\rho(\mathbf{r}) = \sum_{k=1}^{M'} f_k(\mathbf{r} - \mathbf{a}_k) \tag{57}$$

where $\mathcal{A} := \{\mathbf{a}_k\}_{1 \leq k \leq M'}$ is a set of points of the physical space $\mathbb{R}^3$, and $f_k : \mathbb{R}^3 \to \mathbb{R}^3$ functions for which the multipoles

$$(Q_{\ell m}^k) = K_\ell^m \int_{\mathbb{R}^3} \mathcal{R}_\ell^m(\mathbf{r}) f_k(\mathbf{r}) d\mathbf{r}, \tag{58}$$

with $\mathcal{R}_\ell^m(\mathbf{r}) = |\mathbf{r}|^\ell Y_\ell^m \left(\frac{\mathbf{r}}{|\mathbf{r}|}\right)$ (real solid harmonics using $L^2(\mathbb{S}^2)$-orthonormal real spherical harmonics), are easily computable. The coefficients K_ℓ^m are real normalization coefficients (for instance, $K_0^0 = \sqrt{4\pi}$ and $K_1^m = \frac{\sqrt{4\pi}}{3}$).

The purpose of DMA methods is to redistribute the multipoles $Q_{\ell m}^k$ carried by the points $\mathbf{a}_k$ to multipoles $\widetilde{Q}_{\ell m}^j$ carried by some chosen points $\mathbf{S}_j$, called DMA expansion sites. The set $\mathcal{S} = \{\mathbf{S}_j\}_{1 \leq j \leq J}$ can coincide with the set of atomic positions, but can include other points. For instance, Mulliken population analysis defines overlap populations (that can be attributed to the bonds and allocated to bond centers), and the DMA can also include charge contributions from *e.g.* bond centers and from many other points.

The rationale of the distributed multipole analysis is thus the following:

1. first, the multipole moments $(Q_{\ell m}^k)$ are computed. The latter will be coined multipole moments *at natural centers* in the following. In the case of an electronic density computed in a basis of Gaussian Type Orbitals (GTOs) – as used in most quantum chemistry applications –, there is an analytical expression for these elementary moments, allowing for a very fast computation;

2. then, these multipoles are redistributed to the DMA expansion sites $\mathbf{S}_j$, giving rise to the redistributed multipoles $\widetilde{Q}_{\ell m}^j$ carried by the sites $\mathbf{S}_j$.

These two aspects are discussed in the following two sections.

4.1 Multipole moments with respect to natural centers

We recall here the procedure for rewriting as (57) an electronic density ρ issued from a quantum chemistry calculation in a Gaussian type orbital (GTO) basis set, and computing the corresponding multipoles $Q^k_{\ell m}$. First, ρ is decomposed as

$$\rho(\mathbf{r}) = \sum_{\mu,\nu=1}^{N_b} P_{\mu\nu}\, \chi_\mu(\mathbf{r}-\mathbf{R}_{a_\mu})\chi_\nu(\mathbf{r}-\mathbf{R}_{a_\nu}), \tag{59}$$

where $(\chi_\mu)_{1\leq\mu\leq N_b}$ is a basis of *primitive* Gaussian polynomial functions, a_μ the label of the nucleus χ_μ is attached to, and $(P_{\mu\nu})_{1\leq\mu,\nu\leq N_b}$ the density matrix in this basis. The matrix $(P_{\mu\nu})_{1\leq\mu,\nu\leq N_b}$ is easily obtained from the density matrix in the *contracted* GTO basis set used for the computation of ρ by a suitable linear transform. Each χ_μ is of the form

$$\chi_\mu(\mathbf{r}) = N_\mu \mathcal{R}^{m_\mu}_{\ell_\mu}(\mathbf{r}) e^{-\zeta_\mu|\mathbf{r}|^2}, \tag{60}$$

with ℓ_μ and m_μ its angular momentum quantum numbers, ζ_μ the Gaussian exponent controlling its decay rate, and N_μ a normalization constant. Since

$$e^{-\zeta_\mu\left|\mathbf{r}-\mathbf{R}_{a_\mu}\right|^2} e^{-\zeta_\nu|\mathbf{r}-\mathbf{R}_{a_\nu}|^2} = K_{\mu\nu} e^{-(\zeta_\mu+\zeta_\mu)|\mathbf{r}-\mathbf{R}_{\mu\nu}|^2}, \tag{61}$$

with $K_{\mu\nu} = e^{-\frac{\zeta_\mu\zeta_\mu}{\zeta_\mu+\zeta_\nu}\left|\mathbf{R}_{a_\mu}-\mathbf{R}_{a_\nu}\right|^2}$, and $\mathbf{R}_{\mu\nu} = \frac{\zeta_\mu\mathbf{R}_{a_\mu}+\zeta_\nu\mathbf{R}_{a_\nu}}{\zeta_\mu+\zeta_\nu}$, it holds that

$$\chi_\mu(\mathbf{r}-\mathbf{R}_{a_\mu})\chi_j(\mathbf{r}-\mathbf{R}_{a_\nu}) = K_{\mu\nu}N_\mu N_\nu\, \mathcal{R}^{m_\mu}_{\ell_\mu}(\mathbf{r}-\mathbf{R}_{a_\mu})\, \mathcal{R}^{m_\nu}_{\ell_\nu}(\mathbf{r}-\mathbf{R}_{a_\nu})\, e^{-(\zeta_\mu+\zeta_\mu)|\mathbf{r}-\mathbf{R}_{\mu\nu}|^2}, \tag{62}$$

so that the electronic density given by (59) can be written as (57), where the functions $f_k : \mathbb{R}^3 \to \mathbb{R}$ are Gaussian polynomials of the form (62), and $\mathcal{A} := \{\mathbf{a}_k\}_{1\leq k\leq M'}$ a set of points in the physical space, (strictly) containing the positions of the nuclei.

The key quantities for the calculation of the distributed multipole moments are thus the multipole moments with respect to natural expansion centers $\mathbf{R}_{\mu\nu}$:

$$Q^{\mu\nu}_{\ell m} = P_{\mu\nu}K^m_\ell \int_{\mathbb{R}^3} \mathcal{R}^m_\ell(\mathbf{r})\chi_\mu(\mathbf{r}-\mathbf{R}_\mu+\mathbf{R}_{\mu\nu})\chi_\nu(\mathbf{r}-\mathbf{R}_\nu+\mathbf{R}_{\mu\nu})d\mathbf{r}. \tag{63}$$

Given the expression (62), the series of multipole moments $\{Q^{\mu\nu}_{\ell m}\}_{\ell\in\mathbb{N},m=-\ell,\ldots,\ell}$ is finite and can be easily evaluated numerically, by expanding the product of the two solid harmonics in (62) as a weighted sum of homogeneous polynomials, multiplying it with the solid harmonic $\mathcal{R}^m_\ell(\mathbf{r})$ (homogeneous polynomial) and finally using the series of tabulated integrals:

$$I_q = \int_0^{+\infty} u^q e^{-u^2} du = \frac{\Gamma\left(\frac{q+1}{2}\right)}{2}, \tag{64}$$

for all needed q integer values, where Γ is the Euler Gamma function.

4.2 Redistribution formula

Once multipole moments at natural centers are computed, a lower number of expansion sites (than the total number of overlap centers *i.e.* $|\mathcal{A}|$) may be desired to perform the distributed multipole expansion, as the number of overlap centers grows quadratically with the number of atoms or basis size. For instance, only nuclei positions may be retained as final centers (e.g.

to use these local multipole moments in polarizable force fields such as AMOEBA [50]), or all nuclei and bond centers [52, 60]. The multipole moments at natural centers, which are easy to compute (see above), have thus to be redistributed to the chosen final expansion sites. We recall that the set of final expansion sites is a choice taken by the user.

This requires two ingredients. First, the introduction of redistribution weights in order to distribute the multipoles from natural sites not belonging to the set of final sites. Second, a transformation of a multipole corresponding to one expansion center to a multipole corresponding to a different expansion center which is a linear transformation that is known in the Fast Multipole Method (FMM) framework as $M2M$-operator [17] (used here in the context of real-valued multipolar coefficients).

We first focus on the latter. We briefly present the resulting formula, and refer to [8] for a detailed derivation:

$$Q_{\ell m}^{\mu\nu\to j} = \frac{\Re\left(\left(Q_{\ell,-m}^{\mu\nu\to j}\right)^{\mathbb{C}}\right) + (-1)^m \Re\left(\left(Q_{\ell m}^{\mu\nu\to j}\right)^{\mathbb{C}}\right)}{\sqrt{2}} \quad \text{if } m > 0, \tag{65}$$

$$Q_{\ell m}^{\mu\nu\to j} = \frac{(-1)^{|m|}\Im\left(\left(Q_{\ell,-m}^{\mu\nu\to j}\right)^{\mathbb{C}}\right) - \Im\left(\left(Q_{\ell m}^{\mu\nu\to j}\right)^{\mathbb{C}}\right)}{\sqrt{2}} \quad \text{if } m < 0, \tag{66}$$

$$Q_{\ell,0}^{\mu\nu\to j} = \Re\left(\left(Q_{\ell,0}^{\mu\nu\to j}\right)^{\mathbb{C}}\right), \tag{67}$$

where the complex $M2M$-operator is given by

$$\left(Q_{\ell m}^{\mu\nu\to j}\right)^{\mathbb{C}} = \sum_{\ell'=0}^{\ell} \sum_{m'=-\ell'}^{\ell'} c_{\ell' m'}^{\ell m} \left|\mathbf{R}_{\mu\nu} - \mathbf{S}_j\right|^{\ell-\ell'} \mathcal{Y}_{\ell-\ell'}^{m-m'} \left(\frac{\mathbf{R}_{\mu\nu} - \mathbf{S}_j}{|\mathbf{R}_{\mu\nu} - \mathbf{S}_j|}\right) \left(Q_{\ell',m'}^{\mu\nu}\right)^{\mathbb{C}}, \tag{68}$$

where $\mathcal{Y}_\ell^m$ denote the complex ($L^2(\mathbb{S}^2)$-orthonormal) spherical harmonics and $c_{\ell',m'}^{\ell m} = \sqrt{\binom{\ell+m}{\ell'+m'}\binom{\ell-m}{\ell'-m'}}$ a normalization factor. Complex arithmetic operations are used since it simplifies the presentation of the formulae. Finally, the complex multipole moments are simply given by

$$\left(Q_{\ell m}^{\mu\nu}\right)^{\mathbb{C}} = P_{\mu\nu} K_\ell^m \int_{\mathbb{R}^3} |\mathbf{r}|^\ell \mathcal{Y}_\ell^m \left(\frac{\mathbf{r}}{|\mathbf{r}|}\right) \chi_\mu(\mathbf{r} - \mathbf{R}_\mu + \mathbf{R}_{\mu\nu}) \chi_\nu(\mathbf{r} - \mathbf{R}_\nu + \mathbf{R}_{\mu\nu}) d\mathbf{r}. \tag{69}$$

The multipole moments are redistributed to the final expansion sites $\mathbf{S}_j$ using a formula of the form

$$\widetilde{Q}_{\ell m}^j = \sum_{(\mu,\nu)} C_{\mu\nu\to j} \left(Q_{\ell m}^{\mu\nu\to j}\right)^{\mathbb{R}}, \tag{70}$$

where the $C_{\mu\nu\to j}$'s are *user-specified* redistribution weights, representing the part of the multipolar distribution centered at $\mathbf{R}_{\mu\nu}$ allocated to the expansion site $\mathbf{S}_j$. It holds

$$C_{\mu\nu\to j} \geq 0, \quad \sum_{j=1}^{p} C_{\mu\nu\to j} = 1, \quad C_{\mu\nu\to j} = 1 \text{ if } \mathbf{S}_j = \mathbf{R}_{\mu\nu}.$$

In particular, if $\mathbf{R}_{\mu\nu}$ is one of the final expansion sites, then $Q_{\ell m}^{\mu\nu}$ is fully allocated to this site. As a consequence, $Q_{\ell m}^{\mu\nu\to j}$ needs to be computed only if $\mathbf{R}_{\mu\nu}$ is *not* one of the final expansion sites. It is up to the user to choose both the final expansion sites and the values of the weight

coefficients $C_{\mu\nu\to j}$. A complete flexibility is allowed in our DMA code [9]. For instance, following Stone's *neighbor-takes-it-all* strategy implemented in GDMA code [53] would amount to take:

$$C_{\mu\nu\to j} = \begin{cases} 1 \text{ if } \mathbf{S}_j \text{ is the final expansion site nearest to } \mathbf{R}_{\mu\nu}, \\ 0 \text{ otherwise,} \end{cases} \tag{71}$$

or $\frac{1}{q}$ for each of the q final sites equally close to $\mathbf{R}_{\mu\nu}$. This strategy means that multipole moments at natural centers are only redistributed *locally*, at the nearest neighbor final site. An obvious drawback of this approach is that the redistributed moments $\widetilde{Q}^j_{\ell m}$ are clearly not differentiable, nor even continuous, with respect to the nuclear coordinates. We have also implemented, and provide as a possible user-choice, the more balanced rule of Vigne-Maeder *et al.* [60], which amounts to choose the following weight coefficients:

$$C_{\mu\nu\to j} = \frac{\frac{1}{|\mathbf{S}_j - \mathbf{R}_{\mu\nu}|}}{\sum_k \frac{1}{|\mathbf{S}_k - \mathbf{R}_{\mu\nu}|}}, \tag{72}$$

where the sum in the denominator runs over the p chosen final DMA sites. This time, multipole moments at natural centers that are overlooked in the set of DMA expansion sites are redistributed to *all* final DMA sites $\{\mathbf{S}_j\}_j$, with larger contributions to nearer sites.

4.3 Multipolar expansion of the electrostatic potential

Once the local distributed multipole moments are computed, the electrostatic potential generated by the charge density distribution (that the latter summarize) can be evaluated by a simple sum, instead of a three-dimensional integral needed to evaluate the exact, quantum potential:

$$V(\mathbf{r}) = \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'.$$

The multipolar expansion of the total potential with respect to the expansion centers $\mathbf{R}_{\mu\nu}$ then writes:

$$V_{\mathrm{M}}(\mathbf{r}) = \sum_{\mu,\nu=1}^{N_b} \left(\sum_{\ell=0}^{+\infty} \frac{4\pi}{(2\ell+1)} \sum_{m=-\ell}^{\ell} \frac{Q^{\mu\nu}_{\ell m} Y^m_\ell \left(\frac{\mathbf{r}-\mathbf{R}_{\mu\nu}}{|\mathbf{r}-\mathbf{R}_{\mu\nu}|} \right)}{K^m_\ell \, |\mathbf{r} - \mathbf{R}_{\mu\nu}|^{\ell+1}} \right), \tag{73}$$

where the sum over ℓ is actually finite and terminates for $\ell = \ell_\mu + \ell_\nu$, the sum of the degrees of the two real solid harmonics involved in the primitive GTOs χ_μ and χ_ν respectively. The multipolar expansion V_{M} is an excellent approximation of V far enough from the molecule, and a very poor approximation in the region of space where the electronic density is large. In the intermediate region, it is valid *up to a penetration term* [32, 43].

After redistribution of the multipole moments to the DMA final expansion centers, the potential generated by the DMA distributed local multipole moments writes:

$$V^{\mathrm{DMA}}_{\mathrm{M}}(\mathbf{r}) = \sum_{j=1}^{J} \left(\sum_{\ell=0}^{+\infty} \frac{4\pi}{(2l+1)} \sum_{m=-\ell}^{\ell} \frac{\widetilde{Q}^j_{\ell m} Y^m_\ell \left(\frac{\mathbf{r}-\mathbf{S}_j}{|\mathbf{r}-\mathbf{S}_j|} \right)}{K^m_\ell \, |\mathbf{r} - \mathbf{S}_j|^{\ell+1}} \right) \tag{74}$$

where the total multipole moments $(\widetilde{Q}^j_{\ell m})_{1\leq j\leq J}$ with respect to final expansion centers $\mathbf{S}_j$ include the contributions of multipole moments naturally centered at $\mathbf{S}_j$ (if any), as well as of the reallocated (*e.g.* neighboring) multipoles – see Eq. (70). The series of multipole moments $\left(\widetilde{Q}^j_{\ell m}\right)_{l,m}$ with origin $\mathbf{S}_j$ is in general infinite, and must be truncated in practice. For this

reason, $V_{\mathrm{M}}^{\mathrm{DMA}}$ is always an approximation of V_{M} as soon as one of the natural centers $\mathbf{R}_{\mu\nu}$ is not kept as a final DMA expansion site, even away from the numerical support of the density ρ.

Let us finally mention that the new version of DMA [53] can be thought as a mix between this DMA basis-set oriented method and real-space oriented (partitionning) methods, as it associates a redistribution of the density matrix contributions for some (the least diffuse) pairwise products of GTOs with a real-space integration for the contributions of the others (the most diffuse) pairwise products of GTOs, thus allowing to stabilize local multipole moments in the limit of a very large Atomic Orbitals basis for the molecular density ρ.

5 Numerical results

We have implemented the four methods of the ISA family described in Section 2, namely ISA, GISA, L-ISA and MB-ISA, in Python language with a specific focus on accurate quadrature schemes for diatomic molecules. This serves as a proof-of-concept to highlight the theoretical results obtained in this article and can by far not be considered as an exhaustive computational study.

The Distributed Multipole Analysis (DMA) method (Section 4), has also been implemented in Python language in a new, more modular package, allowing for user-specified redistribution strategies (see Eq.(70) and e.g. the two possible choices of coefficients (71) and (72)). Our code has been validated by comparing the results of local multipole moments (up to hexadecapoles) obtained on test molecules (H_2O, NH_3, CH_4, benzene, and ions such as ClO^-) with those obtained by the GDMA code [53], using Stone *neighbor-takes-it-all* redistribution strategy (equation 71). Several basis sets were tested for each molecule. We obtained identical distributed (atomic) local multipole moments up to hexadecapoles, up to numerical accuracy.

All these codes are built upon the functionalities of `cclib` library [40], which allows to parse in a generic way the output of most quantum chemistry packages (14 different packages are supported by this library), without any specific parsing required by the user. This new feature enables an easy comparison of DMA or ISA partitionning results among the most popular quantum chemistry codes, providing more interoperability. The parsing by the existing Stone's GDMA code [53] for DMA was indeed only implemented for Gaussian [19] or Psi4 [34] outputs, and similarly for the existing implementations of the ISA methods in Horton code [58].

5.1 Implementational details

Solver. The ISA AIM decomposition has been computed using two different methods: (i) the fixed-point algorithm described in Section 2.5.4, and (ii) an alternative scheme based on the Newton method applied to a suitable formulation of the Euler-Lagrange equations of the problem (see [8] for further details). Both methods yielded the same results (up to numerical accuracy) in our test cases.

Discretization and numerical quadrature. The AIM methods rely either on a real-space discretization, or on a basis-space representation of the functions spanning the linear or non-linear approximation space.

As already detailed in Section 2, the GISA, L-ISA and MB-ISA methods rely on basis-set representations of pro-atomic densities, through linear combination of normalized Gaussians (for GISA and L-ISA) or Slater functions (for MB-ISA). For GISA and L-ISA, only the coefficients of the linear combination are optimized. For MB-ISA, the coefficients of the expansion *and* the exponents involved in the Slater basis functions are both optimized simultaneously. Typically,

a small number of shells (e.g. 4 for H and 6 for C, O and N [57]) are assigned to each atom in GISA and MB-ISA [59] methods.

On the other hand, the original ISA algorithm introduced by Lillestolen *et al.* [30, 31] (Section 2.5.4) relies on a real space radial and angular discretization by point-evaluations.

In the family of ISA-methods, several integrals arise both when computing the total mass N_a of the atomic densities (imposed as a constraint in the GISA, L-ISA and MB-ISA schemes), the first-order moments of the atomic densities (imposed as a constraint in the MB-ISA scheme), or when computing spherical averages in the original ISA method. This requires a numerical quadrature which is given by a tensor product of a radial and spherical grid. The radial grid can be e.g. an 1D integration grid based on N_r Gauss quadrature points extending up to some maximal R_max, while the angular grid can be e.g. a Lebedev grid [29] made of N_s points on the unit sphere $\mathbb{S}^2$ (with their associated weights). For each atom a, the discretization points are defined as follows:

$$\forall a \in [\![1, M]\!],\ \forall i \in [\![1, N_\mathrm{r}]\!],\ \forall j \in [\![1, N_\mathrm{s}]\!], \qquad \mathbf{r}^a_{i,j} := \mathbf{R}_a + r^a_i \boldsymbol{\sigma}_j, \tag{75}$$

denoting $[\![1, M]\!] = \{1, 2, \ldots, M\}$ and where $\mathbf{R}_a$ is the position of atom a, $(r^a_i)_{i\in[\![1,N_\mathrm{r}]\!]}$ are the (atom-dependent) radial grid points (which can be e.g. equally-spaced or logarithmic), and $(\boldsymbol{\sigma}_j)_{j\in[\![1,N_\mathrm{s}]\!]}$ are the Lebedev grid points grids used for the quadrature on the unit sphere. The molecular density ρ is precomputed at these points $\mathbf{r}^a_{i,j}$, yielding tabulated values $\rho^a_{i,j} = \rho(\mathbf{r}^a_{i,j})$ used throughout the algorithm repeatedly.

In the special case of symmetric systems, such as for some diatomic molecules (associated to an electronic density invariant by any rotation around the axis of the molecule), the spherical average on the unit sphere $\mathbb{S}^2$ can be replaced by a simple one-dimensional integration on the θ variable using N_s points, e.g. using a Gauss-Legendre quadrature by aligning the atoms on the z-axis. All integrals on the radial variable r are computed likewise.

Interpolation. In the ISA-method, where (pro-)atomic densities are known only on a real space grid, we use a piecewise linear (sometimes referred to $\mathbb{P}^1$) interpolation between the nearest neighbors. We use such an interpolation for instance to evaluate the pro-atomic density ρ^0_a of atom a at a radial discretization value r^b_i associated to a near atom b. Indeed, only the values $\rho^0_a(r^a_i)$ for $i \in [1, N_\mathrm{r}]$, at the radial grid point r^a_i, are known. On the contrary, for the GISA, L-ISA and MB-ISA schemes, the expressions of the pro-atomic densities are (by construction) analytical due to the expansion on Gaussian or Slater basis functions and no interpolation is thus needed for these methods.

Optimization. For the resolution of the GISA problem, in particular for the minimization of the quadratic functional under linear (equality and inequality) constraints (45), we use standard routines for quadratic programming problems [20, 21] implemented in the QUADPROG Python function. For solving the convex minimization (38) arising in L-ISA, we use the CVXOPT Python package [2].

Parameterization. For GISA and L-ISA methods, the atomic shell exponents $(\alpha_{a,k})$, with $a \in [\![1, M]\!]$, $k \in [\![1, m_{z_a}]\!]$, are fixed once and for all. The solutions of the GISA and L-ISA schemes thus *a priori* depend on the values of these exponents. For C, H, N and O atoms, we use the values of the exponents provided by Verstraelen *et al.* for the GISA method [57], obtained by fitting to the densities of isolated (neutral or ionized) atoms, so that the GISA Gaussian basis represents accurately enough all the pro-atomic densities associated to the quantum calculations on both isolated neutral atoms and ions. For other atoms, for which no values of the exponents are available, we apply the empirical formula provided by Verstraelen *et al.* in the MB-ISA

paper [59]:

$$\forall k \in [\![1, m_{z_a}]\!], \qquad \alpha_{a,k} = \frac{2 z_a^{1-\frac{k-1}{m_{z_a}-1}}}{a_0}, \tag{76}$$

where a_0 is the value of 1 Bohr in Angström and m_{z_a} the number of shells (derived from the atomic number z_a). For MB-ISA, the exponents $(\alpha_{a,k})$, with $a \in [\![1, M]\!]$, $k \in [\![1, m_{z_a}]\!]$, of the Slater functions used to expand the pro-atomic densities are optimized instead of attributing fixed values. As initial guess for these exponents, we simply take the values of the GISA (or L-ISA) exponents detailed above. These initial values of the exponents appear numerically not to matter (as regards the final fixed-point solution) in the MB-ISA scheme, which optimizes both the weight and the Slater exponent of every pro-atomic shell (see Section 2.7). Note that some choices of exponents lead to an ill-conditionned local overlap matrix $\mathbf{S}_{z_a}$, which can prove problematic in particular in the context of the GISA method (see Section 2.6), where a linear system involving this matrix has to be solved.

For the expansion coefficients $(c_{a,k})$, with $a \in [\![1, M]\!]$, $k \in [\![1, m_{z_a}]\!]$, which are optimized in the GISA, L-ISA and MB-ISA methods, several initial guesses that satisfy the property that the initial pro-atomic mass $N_a^{(0)} = \sum_{k=1}^{m_{z_a}} c_{a,k}$ equals the atomic number z_a, are possible, such as

$$\forall k \in [\![1, m_{z_a}]\!], \qquad c_{a,k}^{(0)} = \frac{z_a}{m_{z_a}}$$

for example. Other less-balanced initial guesses, e.g. putting all the initial weight on a specific Gaussian (or Slater) basis function:

$$\forall k \in [\![1, m_{z_a}]\!], \qquad c_{a,k}^{(0)} = \delta_{k,k_0} z_a,$$

with $k_0 \in [\![1, m_{z_a}]\!]$ and where $\delta_{i,j}$ denotes the Kronecker delta, are also possible.

Uniqueness and dependency on initial condition. For the GISA method, there is no guarantee of the uniqueness of the solution for the expansion coefficients, and local minima are possible (i.e. a dependence of the computed solution on the initial guess for the expansion coefficients). An example found numerically for a very simple test density is provided in the Appendix. On the contrary, for L-ISA, the optimal expansion coefficients $\left(c_{a,k}^{\text{opt}}\right)$ does not depend on the initial guess $\left(c_{a,k}^{(0)}\right)$ due to the strict convexity. For MB-ISA, there is no guarantee of the uniqueness of the fixed-point solution, so that the numerically computed solution may depend on the initial guess. However, numerically, the MB-ISA solution appears quite robust to the guess (both in terms of expansion coefficients $c_{a,k}$ and of exponents $\alpha_{a,k}$), and no local minimum has been unveiled in this study on diatomic systems.

Validation. Our implementation of the different existing ISA partitionning schemes has been validated on some test diatomic molecules, for several basis sets, by comparison to the Horton code [58], specifically for the ISA and MB-ISA scheme. Both implementations (ours and Horton's) yielded the same local multipole moments (up to quadrupoles) up to numerical accuracy on the tested molecules. For GISA and L-ISA, our implementation was validated based on custom test atomic densities made of a sum of one or several normalized Gaussians, centered at both atomic positions (diatomic molecules only).

5.2 Comparison of the different ISA schemes

In this section, we compare typical pro-atomic densities profiles and the associated atomic moments (up to second order) for diatomic molecules or ions, at the equilibrium inter-atomic distance for Hartree-Fock densities obtained with GAMESS [48].

CO molecule.
Figure 1 represents the numerically computed $r \mapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ profiles solutions of GISA, L-ISA, MB-ISA and ISA for the carbon atom in the CO molecule, using a molecular density ρ computed at the RHF/aug-cc-pVDZ level of theory. Qualitatively similar profiles are obtained for the oxygen atom. Note that the quantity $4\pi r^2 \rho_a^0(r)$ represents the contribution of the atomic density ρ_a (whose spherical average equals ρ_a^0) to the atomic charge $N_a = \int_0^{+\infty} 4\pi r^2 \rho_a^0(r) dr$ at a radial distance r.

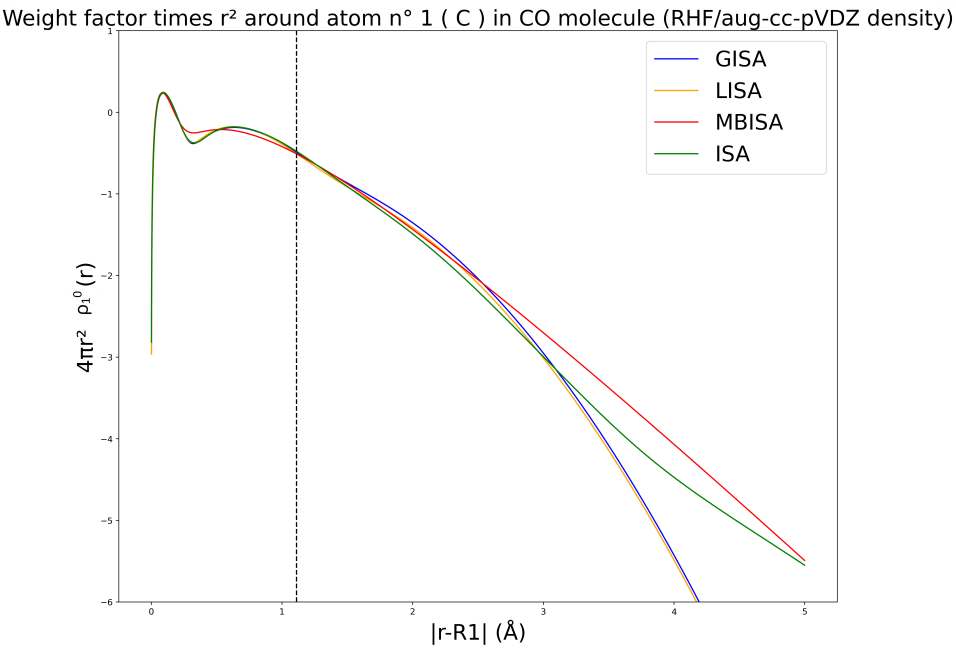

Figure 1: Profiles $r \longmapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ for the carbon atom in CO (1.131 Å interatomic distance, RHF/aug-cc-pVDZ level of theory), obtained by GISA, L-ISA, and MB-ISA and ISA methods. $N_\mathrm{r} = 320$ Gauss-Legendre radial (up to $\mathrm{R_{max}}$=8 Å) and $N_\mathrm{s} = 150$ Gauss-Legendre angular points were used for both the quadratures and the ISA discretization. The black vertical dashed line indicates the position of the oxygen atom.

The GISA and L-ISA profiles look very similar but lead to slightly different atomic multipole moments (see Table 1). The GISA, L-ISA and MB-ISA pro-atomic densities are (visually) very close to the ISA reference near the first maximum of the curve – corresponding to the core electrons (1s) of the carbon atom. Near the second maximum – corresponding to the 2s electrons shell – the GISA and L-ISA profiles are also very close to the ISA reference profile, while the MB-ISA profile features a less pronounced local maximum. The GISA and L-ISA profiles then differ quite notably from the MB-ISA and ISA reference profile in the asymptotic regime.

This proximity of GISA and L-ISA to the ISA reference profile seems rather due to the choice of optimized exponents $\alpha_{a,k}$ for C and O atoms in GISA and L-ISA from Ref. [57]. For the

N_2 molecule (also studied, results not reported), the results are indeed qualitatively similar to those obtained for CO. In the case of optimized exponents $\alpha_{a,k}$, the GISA and L-ISA Gaussian expansions of the pro-atomic densities variational space perform well to capture correctly the short-range features of ρ_a^0. However, this (visual) proximity of GISA and (quantitative) proximity of L-ISA to the reference ISA profile vanishes if the (fixed) exponents $\alpha_{a,k}$ used in GISA and L-ISA are not optimized for one of the atoms in the molecule. Moreover, in the case of non-optimized exponents, only L-ISA recovers qualitatively the reference ISA profile, but not GISA (see below the case of Cl atom in ClO^- ion, Figure 2, compared to Figure 3 for atom O in the same ion).

Multipole moment (atomic units)	GISA	L-ISA	MB-ISA	ISA
Charge $q_1 = q_C$	0.129	0.182	0.225	0.183
Dipole d_z^1	0.217	0.298	0.325	0.307
Dipole d_z^2	-0.015	0.015	0.079	0.008
$Q_{xx}^1 = Q_{yy}^1$	4.109	3.903	3.835	3.887
$Q_{xx}^2 = Q_{yy}^2$	3.476	3.682	3.751	3.696
Q_{zz}^1	5.100	4.823	4.789	4.786
Q_{zz}^2	3.636	3.806	3.917	3.812

Table 1: Local charges, dipoles and second-order moments carried by C (atom 1) and O (atom 2) in CO (1.131 Å interatomic distance, RHF/aug-cc-pVDZ level of theory) computed by GISA, L-ISA, MB-ISA and ISA methods. 320 Gauss-Legendre radial (up to R_{max}=8 Å) and 150 Gauss-Legendre angular points were used for both the quadratures and for the ISA discretization.

The local multipole moments associated to the different ISA profiles of Figure 1 are reported Table 1. The results of the L-ISA scheme are overall the closest to the ISA reference, followed by MB-ISA and GISA, which appears less satisfactory (and for which the atomic monopole differs by more than 30 % from the reference ISA monopole).

ClO^- ion.

Figures 2 and 3 display the $r \mapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ profiles of GISA, L-ISA, MB-ISA and ISA solutions for the chlorine and oxygen atoms respectively, using a molecular density ρ computed at the RHF/6-31G(d) level of theory. Optimized exponents $\alpha_{a,k}$ for Cl atom have not been reported in the literature to our knowledge. We first tried the empirical rule (76) but finally assigned six shells (in GISA and L-ISA) to the chlorine atom with the same (optimized) exponents as for the oxygen atom, as it yielded slightly better (although not satisfactory, see below) results for the GISA profiles and multipole moments. The solution of the MB-ISA method was found not to depend on this initial choice of exponents for the Cl atom (empirical rule (76) or same exponents as the O atom).

Contrary to the case of neutral molecules (CO, N_2) with optimized exponents for both atoms, the pro-atomic densities profiles of GISA, L-ISA and MB-ISA quantitatively differ from the reference ISA profile, with some common features (*e.g.* three local minima for the L-ISA and ISA reference schemes, *vs.* only two for MB-ISA). The GISA profile ends up with an nonphysical, too large (positive) charge (1.94 a.u.) on Cl atom (-2.84 a.u. on O atom) corresponding to a too quick (Figure 2) or (resp.) too slow (Figure 3) decrease in the Cl (resp. O) atomic density profile, despite a good short-range accordance with the ISA reference profile in the case of the O atom (using its optimized Gaussian exponents [57]).The other GISA atomic multipole moments

are also either much too large or largely differing (more than 100 %) from the ISA reference multipoles. The MB-ISA profile is associated to local multipole moments in fair accordance with the ISA reference multipoles (see Table 2) despite differing qualitatively quite notably from the ISA reference profile, in particular in the asymptotic regime. The latter is a feature rather due to the expansion on Slater functions, leading to a decrease at a smaller rate.

Quite notably, unlike the GISA scheme, the L-ISA method is also associated to local multipole moments in fair accordance with the ISA reference multipoles (Table 2), similarly to the MB-ISA method, although the latter optimizes on the (Slater) exponents while the L-ISA uses fixed (non-optimized) Gaussian exponents. The L-ISA scheme, which has been shown here to be mathematically more grounded than GISA, seems to be more robust with respect to the Gaussian exponent choices and yields more physical (and closer to the ISA reference profile) results, although this assumption would have to be confirmed on larger test sets.

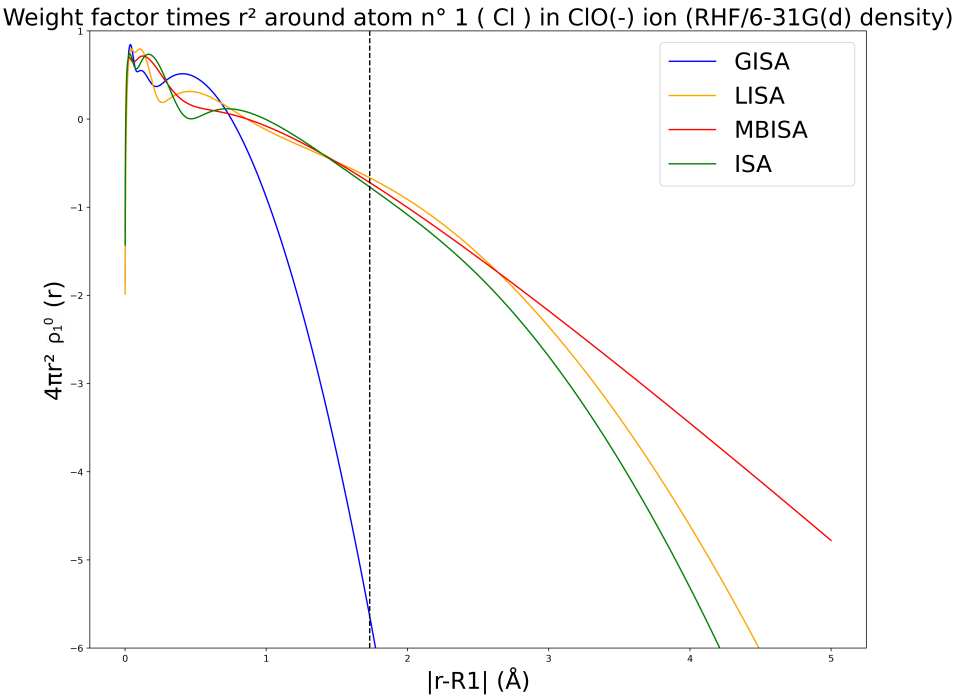

Figure 2: Profiles $r \longmapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ for the chlorine atom in ClO^- ion (1.733 Å interatomic distance, RHF/6-31G(d) level of theory), obtained by GISA, L-ISA, and MB-ISA and ISA methods. $N_\mathrm{r} = 500$ Gauss-Legendre radial (up to $\mathrm{R_{max}}$=8 Å) and $N_\mathrm{s} = 250$ Gauss-Legendre angular points were used for both the quadratures and the ISA discretization. The black vertical dashed line indicates the position of the oxygen atom.

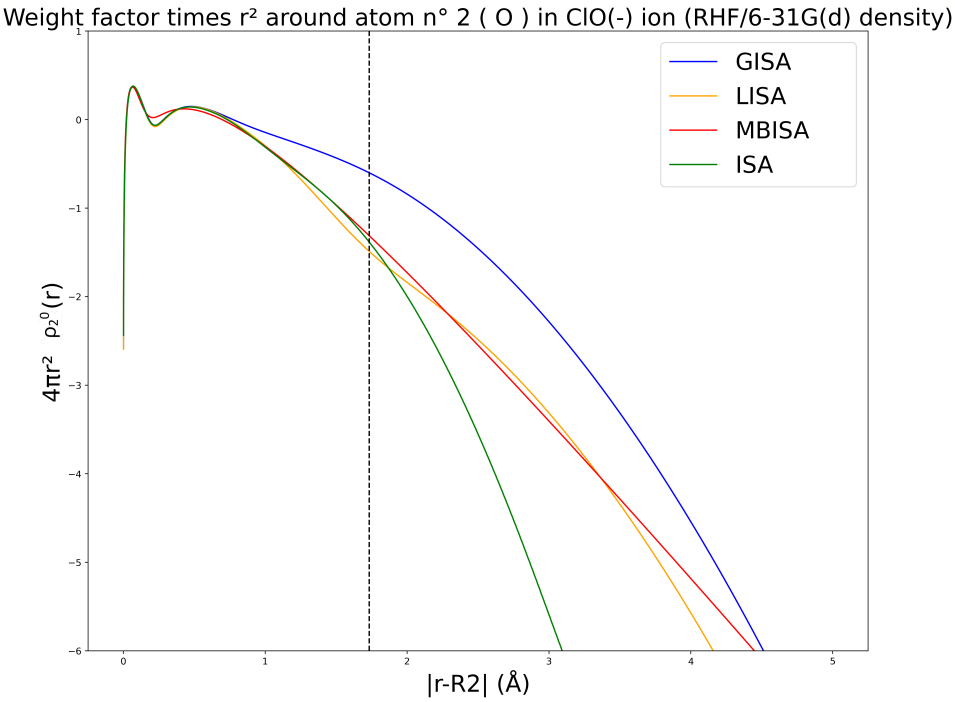

Figure 3: Profiles $r \longmapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ for the oxygen atom of ClO^- ion (1.733 Å interatomic distance, RHF/6-31G(d) level of theory), obtained by GISA, L-ISA, and MB-ISA and ISA methods. $N_\mathrm{r} = 500$ Gauss-Legendre radial (up to $\mathrm{R_{max}}$=8 Å) and $N_\mathrm{s} = 250$ Gauss-Legendre angular points were used for both the quadratures and the ISA discretization. The black vertical dashed line indicates the position of the chlorine atom.

Multipole moment (atomic units)	GISA	L-ISA	MB-ISA	ISA
Charge $q_1 = q_{Cl}$	1.941	-0.368	-0.328	-0.342
Dipole d_z^1	-1.024	-0.117	-0.198	-0.232
Dipole d_z^2	-6.569	0.086	0.033	0.115
$Q_{xx}^1 = Q_{yy}^1$	4.455	10.460	10.333	10.353
$Q_{xx}^2 = Q_{yy}^2$	10.581	4.576	4.703	4.682
Q_{zz}^1	4.176	9.787	9.567	9.364
Q_{zz}^2	28.460	4.026	4.159	3.983

Table 2: Local charges, dipoles and second-order moments carried by Cl (atom 1) and O (atom 2) in ClO^- (1.733 Å interatomic distance, RHF/6-31G(d) level of theory), obtained by GISA, L-ISA, and MB-ISA and ISA methods. $N_\mathrm{r} = 500$ Gauss-Legendre radial (up to $\mathrm{R_{max}}$=8 Å) and $N_\mathrm{s} = 250$ Gauss-Legendre angular points were used for both the quadratures and the ISA discretization.

5.3 Dissociation of diatomic molecules

In this section, we apply the ISA multi-center decomposition schemes implemented in this work to the study of the dissociation of LiH. We consider ground-state densities obtained by Full CI calculations in a small basis sets (computed with PySCF [55, 56]), for different inter-atomic distances, and extract the corresponding one-body density matrices. We then compute the corresponding ISA atomic charges and dipoles. The results are reported in Figure 4.

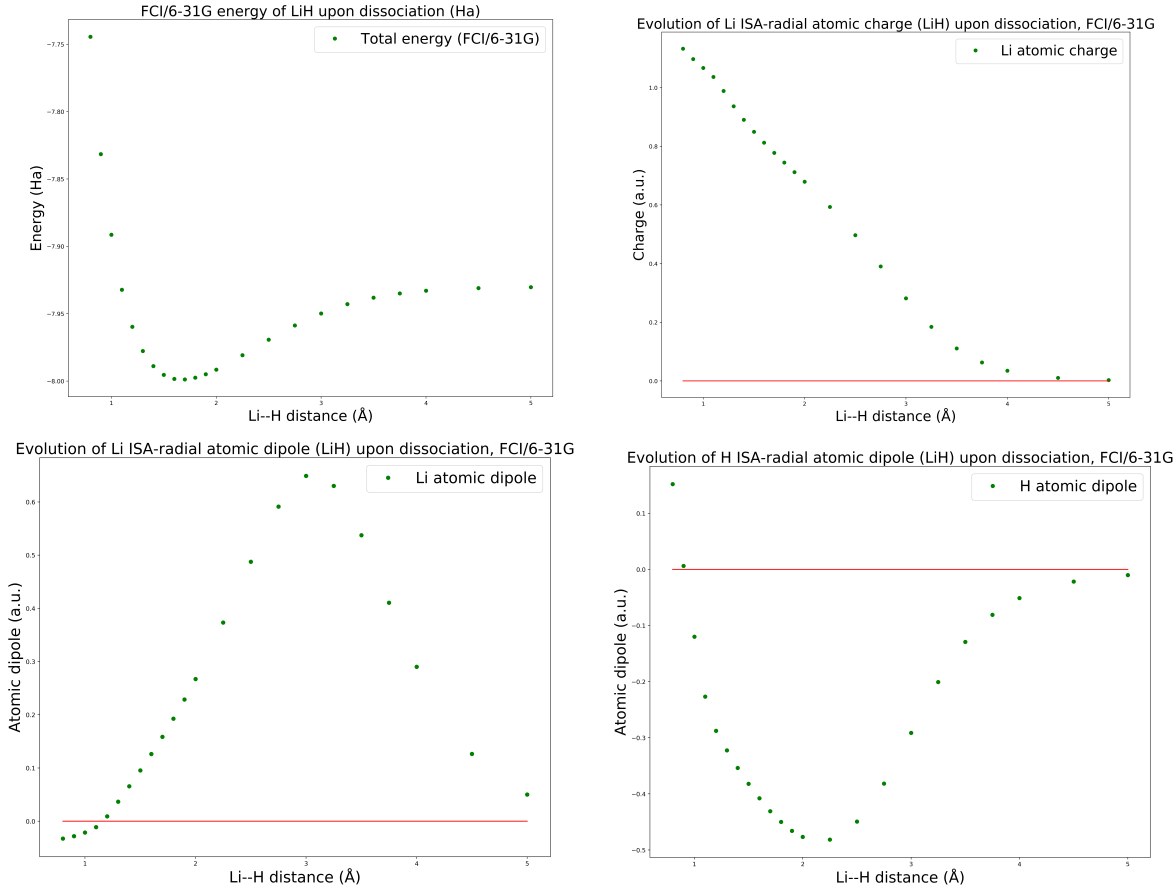

Figure 4: Quantities of interest for LiH as functions of the interatomic distance, with a density obtained at the FCI/6-31G level of theory. Top-left : ground state energy (Ha). Top-right : ISA atomic charge on Li. Bottom: ISA atomic dipoles along the dimer axis on Li (left panel) and H (right panel).

As expected, the ISA atomic charge tends to zero in the dissociation limit, while the Li–H bond is very polarized in the covalent regime (with atomic charges of order 0.80 a.u. at the equilibrium distance 1.70 Å). The ISA Li and H atomic dipoles also tend to zero in the dissociation limit, but in a non-monotonic manner, with a minimum (resp. maximum) for H (resp. Li) atomic dipole at about 2.2 Å (resp. 3.0 Å).

5.4 Basis set convergence of DMA and ISA

In this section, we compare some typical features of the DMA and ISA methods, such as the convergence of local multipole moments with increasing basis sets. Figure 5 displays the evolution of the local atomic charges on H and O atoms of H_2O computed with DMA (using both Stone's and Vigné-Maeder's redistribution strategies) and ISA (computed with Hipart code, now known as Horton [58]), from a molecular density computed at the DFT/PBE0/(aug)-cc-pVXZ (X $\in \{D, T, Q, 5\}$) level of theory [1]. ISA atomic charges stabilize up to a few % for basis sets larger than aug-cc-pVDZ, the DMA atomic charges do not stabilize with increasing basis set size (e.g. four-fold increase of atomic charge on the oxygen atom). This was already known for DMA atomic charges computed with Stone's redistribution strategy and attributed to the presence of diffuse functions in the basis set, and had lead to the hybrid real-space and basis-space GDMA method [53]. We see here that the more balanced Vigné-Maeder redistribution strategy [60] leads to a qualitatively similar non-converging behavior. A similar behavior is observed for atomic dipoles (Figure 6), with ISA dipoles stabilizing up to 10 % beyond aug-cc-pVDZ basis.

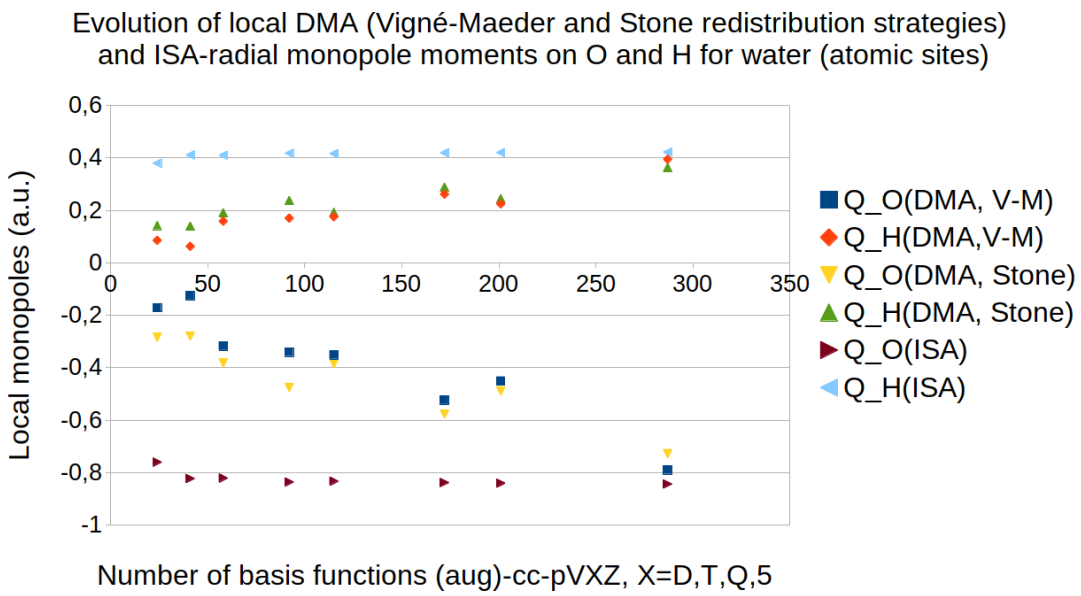

Figure 5: Local atomic charges on hydrogen and oxygen atoms of the water molecule with increasing basis set size, computed from DMA (Stone's and Vigné-Maeder's redistribution strategies) and ISA. The number of basis functions refers to the whole molecule (two hydrogen atoms and oxygen atoms).

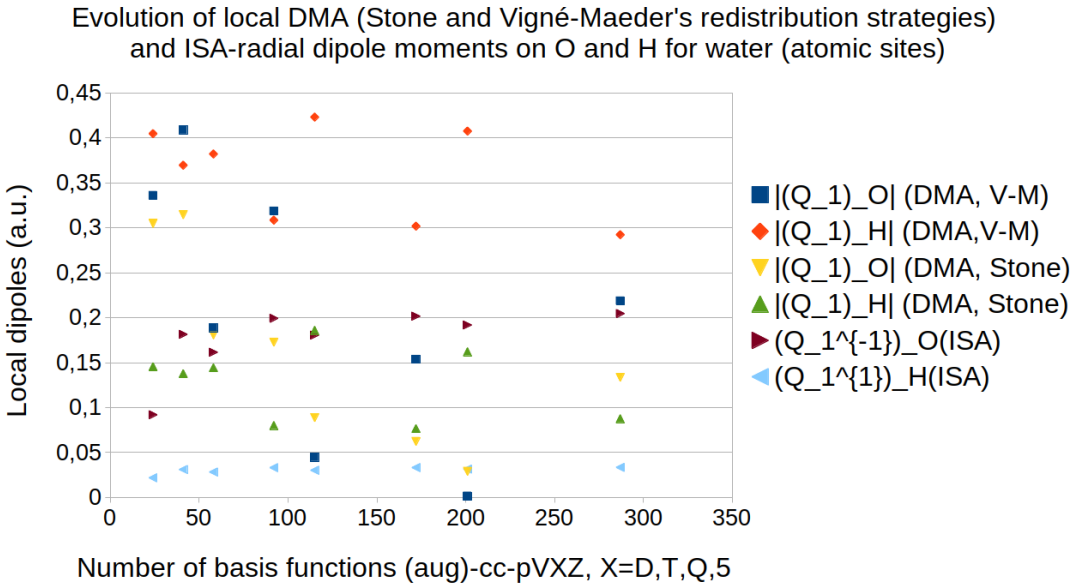

Figure 6: Local atomic dipoles on hydrogen and oxygen atoms of the water molecule with increasing basis set size, computed from DMA (Stone's and Vigné-Maeder's redistribution strategies) and ISA. The number of basis functions refers to the whole molecule (two hydrogen atoms and oxygen atoms).

6 Proofs

Let us first elaborate on the assumptions on ρ and the ρ_a^0's made in the lemmata, propositions and theorems to be proved. As already mentioned previously, electronic densities of molecules and clusters are continuous, positive, functions on $\mathbb{R}^3$, decaying exponentially fast at infinity. We can therefore assume without loss of generality that $\rho \in X_+$, and this implies in particular that $\rho \log \rho \in L^1(\mathbb{R}^3)$. For the same reasons, it makes sense to restrict ourselves to proatom densities ρ_a^0 belonging to X_+. In addition, if the ρ_a's are nonnegative measurable functions satisfying $\sum_{a=1}^M \rho_a(\cdot - \mathbf{R}_a) = \rho$ for some $\rho \in X_+$, then each ρ_a satisfies $0 \leq \rho_a \leq \rho(\cdot + \mathbf{R}_a)$, and therefore belongs to X_+. This justifies *a posteriori* the condition $\boldsymbol{\rho} \in X_+^M$ in the definition (3) of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$.

The following technical remark is important as it allows one to justify several algebraic manipulations in the proofs below.

Remark 10. *Let $\rho \in X_+$. For $\boldsymbol{\rho} = (\rho_a)_{1\leq a\leq M} \in \mathcal{K}_{\rho,\mathbf{R}}$, we have $\rho_a \in X_+$, and therefore $\rho_a \log \rho_a \in L^1(\mathbb{R}^3)$. As a consequence, we have for all $\boldsymbol{\rho} = (\rho_a)_{1\leq a\leq M} \in \mathcal{K}_{\rho,\mathbf{R}}$ and $\boldsymbol{\rho}^0 =$*

$(\rho_a^0)_{1\le a\le M} \in (X_+^r)^M$,

$$\begin{aligned} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho_a}{\rho_a^0}\right) \\ &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \left(\log \rho_a - \log \rho_a^0\right) \qquad \text{(in agreement with (12))} \\ &= \sum_{a=1}^M \int_{\mathbb{R}^3} \left(\underbrace{\rho_a \log \rho_a}_{\in L^1(\mathbb{R}^3)} - \underbrace{(\rho_a \log(\rho_a^0))_+}_{\in L^1(\mathbb{R}^3)\cap L^\infty(\mathbb{R}^3)} + \underbrace{(\rho_a \log(\rho_a^0))_-}_{\ge 0} \right), \end{aligned} \tag{77}$$

where $x_+ = \max(x,0)$ and $x_- = -\min(x,0)$ denote the positive and negative parts of $x \in \mathbb{R}$ ($x_+, x_- \ge 0$ and $x = x_+ - x_-$). In addition, the function $(\rho_a \log(\rho_a^0))_+$ has compact support since ρ_a^0 vanishes at infinity. It follows that $S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) < +\infty$ if and only if $\rho_a \log(\rho_a^0) \in L^1(\mathbb{R}^3)$ for all $1 \le a \le M$, and that we then have (using Fubini theorem)

$$S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log \rho_a - \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log \rho_a^0 = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log \rho_a - \sum_{a=1}^M \int_{\mathbb{R}^3} \langle \rho_a \rangle_s \log \rho_a^0. \tag{78}$$

6.1 Proof of Lemma 2

Let us define

$$\rho_a^*(\boldsymbol{r}) := \left| \begin{array}{ll} \frac{\rho_a^0(\boldsymbol{r})}{\rho^0(\boldsymbol{r}+\mathbf{R}_a)} \rho(\boldsymbol{r}+\mathbf{R}_a) & \text{if } \rho_a^0(\boldsymbol{r}) > 0, \\ 0 & \text{if } \rho_a^0(\boldsymbol{r}) = 0 \end{array} \right. \tag{79}$$

(recall that $\rho^0(\mathbf{r}) := \sum_{a=1}^M \rho_a^0(\mathbf{r}-\mathbf{R}_a)$), and $\boldsymbol{\rho}^* := (\rho_a^*)_{1\le a\le M}$ as an Ansatz for the minimizer of (17). Let

$$B_a^0 := \{\mathbf{r} \in \mathbb{R}^3 \mid \rho_a^0(\boldsymbol{r}) = 0\}, \quad B^0 := \{\mathbf{r} \in \mathbb{R}^3 \mid \rho^0(\boldsymbol{r}) = 0\}, \quad B = \{\mathbf{r} \in \mathbb{R}^3 \mid \rho(\boldsymbol{r}) = 0\}.$$

It follows from the hypothesis (15) (i.e. since $s_{\mathrm{KL}}(\rho|\rho^0) < +\infty$) and the conventions (12) that $B^0 \subset B$. In addition, we have $B^0 = \bigcap_{a=1}^M (\mathbf{R}_a + B_a^0)$, so that if $\mathbf{r} \in B^0$, then for all $1 \le a \le M$, $\rho_a^0(\mathbf{r}-\mathbf{R}_a) = 0$, and therefore, using (79), $\rho_a^*(\mathbf{r}-\mathbf{R}_a) = 0$. Thus, on B^0, there holds $\sum_{a=1}^M \rho_a^*(\cdot - \mathbf{R}_a) = 0 = \rho$ as $B^0 \subset B$. Besides, we have

$$\forall 1 \le a \le M, \quad \forall \mathbf{r} \in \mathbb{R}^3 \setminus B^0, \quad \rho_a^*(\mathbf{r}-\mathbf{R}_a) = \frac{\rho_a^0(\boldsymbol{r}-\mathbf{R}_a)}{\rho^0(\boldsymbol{r})} \rho(\boldsymbol{r}).$$

Thus, there holds that $\sum_{a=1}^M \rho_a^*(\cdot - \mathbf{R}_a) = \rho$ on $\mathbb{R}^3 \setminus B^0$ as well, and therefore on the entire space $\mathbb{R}^3$. Since the ρ_a^*'s are obviously nonnegative, bounded (by $\rho(\mathbf{r}+\mathbf{R}_a)$), integrable, with finite first moments, and vanishing at infinity, it holds that $\boldsymbol{\rho}^* \in \mathcal{K}_{\rho,\mathbf{R}}$. Now, we have

$$\begin{aligned} S(\boldsymbol{\rho}^*|\boldsymbol{\rho}^0) &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a^* \log\left(\frac{\rho_a^*}{\rho_a^0}\right) = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a^*(\cdot - \mathbf{R}_a) \log\left(\frac{\rho_a^*(\cdot - \mathbf{R}_a)}{\rho_a^0(\cdot - \mathbf{R}_a)}\right) \\ &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a^*(\cdot - \mathbf{R}_a) \log\left(\frac{\rho}{\rho^0}\right) = \int_{\mathbb{R}^3} \rho \log\left(\frac{\rho}{\rho^0}\right) = s_{\mathrm{KL}}(\rho|\rho^0) < +\infty. \end{aligned}$$

All these equalities can be rigorously justified by using arguments similar as in Remark 10 and the infimum in (17) therefore has a finite value. For all $\boldsymbol{\rho} = (\rho_a)_{1\le a\le M} \in \mathcal{K}_{\rho,\mathbf{R}}$ such that

$S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) < +\infty$, we have in addition, using the fact that $\rho = \sum_{a=1}^M \rho_a(\cdot - \mathbf{R}_a)$,

$$S(\boldsymbol{\rho}^*|\boldsymbol{\rho}^0) = \int_{\mathbb{R}^3} \rho \log\left(\frac{\rho}{\rho^0}\right) = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a(\cdot - \mathbf{R}_a) \log\left(\frac{\rho}{\rho^0}\right) = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho(\cdot + \mathbf{R}_a)}{\rho^0(\cdot + \mathbf{R}_a)}\right).$$

Therefore, using again arguments as in Remark 10 to justify each equality, we get

$$\begin{aligned} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) - S(\boldsymbol{\rho}^*|\boldsymbol{\rho}^0) &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho_a}{\rho_a^0}\right) - \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho(\cdot + \mathbf{R}_a)}{\rho^0(\cdot + \mathbf{R}_a)}\right) \\ &= \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho_a \rho^0(\cdot + \mathbf{R}_a)}{\rho_a^0 \rho(\cdot + \mathbf{R}_a)}\right) = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a \log\left(\frac{\rho_a}{\rho_a^*}\right) = S(\boldsymbol{\rho}|\boldsymbol{\rho}^*) \geq 0, \end{aligned}$$

for all $\boldsymbol{\rho} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$, which proves that ρ^* is a minimizer of (17). As this problem is strictly convex, it is the unique one.

6.2 Proof of Lemma 5

Obviously, $\langle \rho_a \rangle_s \in \mathcal{K}^0_{\mathrm{ISA}} = X^r_+$ and $\int_{\mathbb{R}^3} \langle \rho_a \rangle_s = \int_{\mathbb{R}^3} \rho_a$, so that $\langle \rho_a \rangle_s$ is in the minimization set of (30). Reasoning as in Remark 10, we obtain that

$$s_{\mathrm{KL}}(\rho_a|\langle \rho_a \rangle_s) = \int_{\mathbb{R}^3} \rho_a \log(\rho_a) - \int_{\mathbb{R}^3} \langle \rho_a \rangle_s \log(\langle \rho_a \rangle_s) < +\infty,$$

and that, for all $\rho_a^0 \in \mathcal{K}^0_{\mathrm{ISA}}$ such that $s_{\mathrm{KL}}(\rho_a|\rho_a^0) < +\infty$, it holds

$$s_{\mathrm{KL}}(\rho_a|\rho_a^0) = \int_{\mathbb{R}^3} \rho_a \log(\rho_a) - \int_{\mathbb{R}^3} \langle \rho_a \rangle_s \log(\rho_a^0).$$

Therefore, for all $\rho_a^0 \in \mathcal{K}^0_{\mathrm{ISA}}$ such that $s_{\mathrm{KL}}(\rho_a|\rho_a^0) < +\infty$, we have

$$\begin{aligned} s_{\mathrm{KL}}(\rho_a|\rho_a^0) &= s_{\mathrm{KL}}(\rho_a|\langle \rho_a \rangle_s) + \int_{\mathbb{R}^3} \langle \rho_a \rangle_s \log(\langle \rho_a \rangle_s) - \int_{\mathbb{R}^3} \langle \rho_a \rangle_s \log(\rho_a^0) \\ &= s_{\mathrm{KL}}(\rho_a|\langle \rho_a \rangle_s) + s_{\mathrm{KL}}(\langle \rho_a \rangle_s|\rho_a^0) \geq s_{\mathrm{KL}}(\rho_a|\langle \rho_a \rangle_s). \end{aligned}$$

This implies that $\langle \rho_a \rangle_s$ is a minimizer of (30) and the infimum is finite. The set $\mathcal{K}^0_{\mathrm{ISA}}$ is a non-empty closed convex subset of the vector space X, and the map $\mathcal{J}_0 := \mathcal{K}^0_{\mathrm{ISA}} \ni \rho_a^0 \mapsto s_{\mathrm{KL}}(\rho_a|\rho_a^0) \in \mathbb{R} \cup \{+\infty\}$ is convex, and strictly convex on the convex set on which it takes finite values. The uniqueness of the solution to (30) is a direct consequence of this strict convexity property.

6.3 L-ISA counterpart of Lemma 5

For $\boldsymbol{\lambda} := (\lambda_a)_{1 \leq a \leq M} \in \mathbb{R}^M$, $\boldsymbol{\mu} := (\mu_a)_{1 \leq a \leq M} \in \mathbb{R}^M$, and $\boldsymbol{\rho} := (\rho_a)_{1 \leq a \leq M} \in X^M$, we denote by $\boldsymbol{\lambda} \odot \boldsymbol{\mu} \in \mathbb{R}^M$ and $\boldsymbol{\lambda} \odot \boldsymbol{\rho} \in X^M$ the Kronecker products defined by

$$[\boldsymbol{\lambda} \odot \boldsymbol{\mu}]_a = \lambda_a \, \mu_a \quad \text{and} \quad [\boldsymbol{\lambda} \odot \boldsymbol{\rho}]_a = \lambda_a \, \rho_a.$$

Lemma 11. *Let $\boldsymbol{\rho} = (\rho_a)_{1 \leq a \in M} \in X_+^M$ such that for all $1 \leq a \leq M$, there exists $\rho_a^0 \in \mathcal{K}^0_{z,\mathrm{L-ISA}}$ such that $\int_{\mathbb{R}^3} \rho_a^0 = \int_{\mathbb{R}^3} \rho_a$ and $s_{\mathrm{KL}}(\rho_a|\rho_a^0) < +\infty$. Then, the problem*

$$\inf_{\boldsymbol{\rho}^0 \in \boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}},\, \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho})} S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) \tag{80}$$

has a unique minimizer $\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho})$. For all $\boldsymbol{\lambda} \in (\mathbb{R}_+)^M$, it holds

$$\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}}(\boldsymbol{\lambda} \odot \boldsymbol{\rho}) = \boldsymbol{\lambda} \odot \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho}). \tag{81}$$

Let us point out that problem (80) can be decomposed as M independent problems of the form

$$\inf_{\rho_a^0 \in \mathcal{K}^0_{z_a,\mathrm{L-ISA}},\, \int_{\mathbb{R}^3} \rho_a^0 = \int_{\mathbb{R}^3} \rho_a} s_{\mathrm{KL}}(\rho_a|\rho_a^0)$$

for all $1 \leq a \leq M$ as in (25).

Proof. The set $\{\boldsymbol{\rho}^0 \in \boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}} \,|\, \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0) = \boldsymbol{\mathcal{N}}(\boldsymbol{\rho})\}$ is non-empty and convex, and is included in a finite-dimensional vector subspace of X. The existence of a solution (80) follows from the compactness of $\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$ (for the strong topology of X^M) and the strong lower-semicontinuity of the functional $\mathcal{J}_0$ introduced in the previous section. The second assertion of the lemma is a consequence of the fact that (i) for all $\boldsymbol{\lambda} \in {\mathbb{R}_+}^M$, $\boldsymbol{\rho} \in X^M$, and $\boldsymbol{\rho}^0 \in \mathcal{K}^0_{\mathrm{L-ISA}}$, $\boldsymbol{\lambda} \odot \boldsymbol{\rho}^0 \in \mathcal{K}^0_{\mathrm{L-ISA}}$, $\boldsymbol{\mathcal{N}}(\boldsymbol{\lambda} \odot \boldsymbol{\rho}^0) = \boldsymbol{\mathcal{N}}(\boldsymbol{\lambda} \odot \boldsymbol{\rho})$ and (ii) $s_{\mathrm{KL}}(\lambda f|\lambda g) = \lambda s_{\mathrm{KL}}(f|g)$ for all $\lambda \in \mathbb{R}_+$ and $f, g \in X_+$. □

6.4 Topological properties of the set of admissible AIM decompositions

The following lemma provides some information about the topology of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ which will be useful for our analysis. Recall that $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)$ is the Banach space of the bounded (signed) Radon measures on $\mathbb{R}^3$ endowed with the total variation norm and that $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)$ is the dual of the Banach space $C_0(\mathbb{R}^3)$ of the real-valued continuous functions going to zero at infinity, and that a sequence $(\mu_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)$ weakly-* converge to some $\mu \in \mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)$ if and only if

$$\forall f \in C_0(\mathbb{R}^3), \quad \int_{\mathbb{R}^3} f \, d\mu_n \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^3} f \, d\mu.$$

Lemma 12. *Let $\rho \in X_+$ and $\mathbf{R} = (\mathbf{R}_1, \ldots, \mathbf{R}_M) \in \mathbb{R}^{3M}$. The set $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ of admissible AIM decompositions defined by (3) is nonempty and convex. It is also bounded and closed for the norm topology of the Banach space X^M, and sequentially compact in the following sense: from any sequence $(\boldsymbol{\rho}^n)_{n\in\mathbb{N}} = ((\rho_a^n)_{1\leq a\leq M})_{n\in\mathbb{N}}$ of elements of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$, one can extract a subsequence $(\boldsymbol{\rho}^{n_k})_{k\in\mathbb{N}}$ converging toward an element $\boldsymbol{\rho}^*$ of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ for the weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, and for the weak topology of $L^p(\mathbb{R}^3)^M$ for any $1 < p < \infty$, and such that*

$$\forall 1 \leq a \leq M, \quad \int_{\mathbb{R}^3} \rho_a^{n_k}(\mathbf{r}) \, d\mathbf{r} \underset{k\to\infty}{\longrightarrow} \int_{\mathbb{R}^3} \rho_a^*(\mathbf{r}) \, d\mathbf{r}, \tag{82}$$

$$\int_{\mathbb{R}^3} |\mathbf{r}| \rho_a^{n_k}(\mathbf{r}) \, d\mathbf{r} \underset{k\to\infty}{\longrightarrow} \int_{\mathbb{R}^3} |\mathbf{r}| \rho_a^*(\mathbf{r}) \, d\mathbf{r}. \tag{83}$$

Proof. Let $(\boldsymbol{\rho}^n)_{n\in\mathbb{N}}$ denote a sequence of elements of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ so that for all $n \in \mathbb{N}$, $\boldsymbol{\rho}^n := (\rho_a^n)_{n\in\mathbb{N}}$. Since for all $n \in \mathbb{N}$ and $1 \leq a \leq M$, $\rho_a^n \geq 0$ and $\rho \geq 0$, the fact that $\sum_{a=1}^M \rho_a^n(\cdot - \mathbf{R}_a) = \rho$ implies that $\|\rho_a^n\|_{L^1(\mathbb{R}^3)} \leq \|\rho\|_{L^1(\mathbb{R}^3)}$ and $\|\rho_a^n\|_{L^\infty(\mathbb{R}^3)} \leq \|\rho\|_{L^\infty(\mathbb{R}^3)}$. Thus, up to the extraction of a subsequence, still denoted by $(\rho_a^n)_{n\in\mathbb{N}}$ for the sake of simplicity, there exists a bounded Radon measure ρ_a^* such that

$$\rho_a^n \underset{n\to\infty}{\rightharpoonup} \rho_a^* \text{ weakly-* in } \mathcal{M}_{\mathrm{b}}(\mathbb{R}^3).$$

In particular, $\rho_a^* \geq 0$ and $\sum_{a=1}^N \rho_a^*(\cdot - \mathbf{R}_a) = \rho$. Since $(1 + |\cdot|)\rho \in L^1(\mathbb{R}^3)$, ρ_a^* is necessarily absolutely continuous with respect to the Lebesgue measure and hence $(1 + |\cdot|)\rho_a^* \in L^1(\mathbb{R}^3)$. Besides, $(\rho_a^n)_{n\in\mathbb{N}}$ being bounded in every $L^p(\mathbb{R}^3)$ for $1 < p \leq \infty$, it holds that, up to the extraction of a subsequence,

$$\rho_a^n \underset{n\to+\infty}{\rightharpoonup} \rho_a^* \text{ weakly-* in } L^\infty(\mathbb{R}^3) \text{ and weakly in } L^p(\mathbb{R}^3), \quad \text{for all } 1 < p < \infty.$$

Thus $\rho^* = (\rho_a^*)_{1\le a\le M} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$, and the sequential compactness properties listed in the lemma hold true. Moreover, let us denote by $N_a^n := \int_{\mathbb{R}^3} \rho_a^n$ and by $N_a^* := \int_{\mathbb{R}^3} \rho_a^*$. It holds that $N_a^* \le \liminf_{n\to+\infty} N_a^n$. The fact that for all $n \in \mathbb{N}^*$, $\sum_{a=1}^M N_a^n = \int_{\mathbb{R}^3} \rho = \sum_{a=1}^M N_a^*$ implies (82). The convergence (83) is obtained in the same way. □

6.5 Proof of Theorem 6

Using the arguments in Remark 4 for ISA, and assumption (51) for L-ISA, we see that the infimum in (50) is finite. Let $\left(\boldsymbol{\rho}^n, \boldsymbol{\rho}^{0,n}\right)_{n\in\mathbb{N}}$ be a minimizing sequence for (50). By construction, the sequence $\left(\boldsymbol{\rho}^n, \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^n)\right)_{n\in\mathbb{N}}$, where

$$\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^n) := \underset{\widetilde{\boldsymbol{\rho}}^0\in\boldsymbol{\mathcal{K}}^0,\, \boldsymbol{\mathcal{N}}(\widetilde{\boldsymbol{\rho}}^0)=\boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^n)}{\arg\min} S(\boldsymbol{\rho}^n|\widetilde{\boldsymbol{\rho}}^0)$$

(we use the notation $\widetilde{\boldsymbol{\rho}}^0$ to avoid a conflict of notation with the first term of the sequence $(\boldsymbol{\rho}^n)_{n\in\mathbb{N}}$), is uniquely defined for ISA and L-ISA by virtue of Lemmata 5 and 11, is also a minimizing sequence to (50). We can therefore assume without loss of generality that $\boldsymbol{\rho}^{0,n} = \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^n)$.

We infer from Lemma 12 that there exists $\boldsymbol{\rho}^* \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ such that, up to extraction of a subsequence, $(\boldsymbol{\rho}^n)_{n\in\mathbb{N}}$ converges to $\boldsymbol{\rho}^*$ for the weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, and for the weak topology of $L^p(\mathbb{R}^3)^M$ for any $1 < p < \infty$, and

$$\forall 1 \le a \le M, \quad N_a^n := \int_{\mathbb{R}^3} \rho_a^n(\mathbf{r})\, dr \underset{k\to\infty}{\longrightarrow} \int_{\mathbb{R}^3} \rho_a^*(\mathbf{r})\, dr =: N_a^*.$$

$\boxed{1}$ We want to show that $\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^n) \rightharpoonup \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^*)$ for the weak-* topology of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$.

(i) Let us first deal with ISA. We have $\rho_a^{0,n} := (\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^n))_a = \langle \rho_a^n \rangle_s$ (spherical average). We therefore have for all $1 \le a \le M$ and all $n \in \mathbb{N}$, $0 \le \rho_a^{0,n} \le \langle \rho(\,\cdot + \mathbf{R}_a)\rangle_s \in X_+$ so that $\rho_a^{0,n} \in X_+$. Reasoning as in the proof of Lemma 12, we obtain the existence of a function $\rho_a^{0,*} \in X_+$ such that

$$\rho_a^{0,n} \underset{n\to+\infty}{\rightharpoonup} \rho_a^{0,*}, \text{ weakly-* in } \mathcal{M}_{\mathrm{b}}(\mathbb{R}^3) \text{ and } L^\infty(\mathbb{R}^3), \text{ and weakly in } L^p(\mathbb{R}^3),$$

for all $1 < p < \infty$. The function $\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}$ is linear and strongly continuous on e.g. $L^2(\mathbb{R}^3)^M$, hence weakly continuous on this space. We therefore have $\boldsymbol{\rho}^{0,n} = \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^n) \rightharpoonup \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^*)$ weakly in $L^2(\mathbb{R}^3)^M$. Since $(\boldsymbol{\rho}^{0,n})_{n\in\mathbb{N}}$ also converges to $\boldsymbol{\rho}^{0,*}$ weakly in $L^2(\mathbb{R}^3)^M$, we obtain that $\boldsymbol{\rho}^{0,*} = G^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^*)$. Thus, $\boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^n) \rightharpoonup \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}}(\boldsymbol{\rho}^*)$ for the weak-* topology of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$.

(ii) Let us now turn to L-ISA. The above argument for ISA heavily relies on the fact that $\boldsymbol{\rho} \mapsto \boldsymbol{G}^0_{\mathcal{K}^0_{\mathrm{ISA}}}(\boldsymbol{\rho})$ is a linear map, which is not the case of the map $\boldsymbol{\rho} \mapsto \boldsymbol{G}^0_{\mathcal{K}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho})$.

We denote by $\overline{\mathcal{K}}^0_{z_a} := \left\{\overline{\rho}_a^0 \in \mathcal{K}^0_{z_a,\mathrm{L-ISA}} \mid \int_{\mathbb{R}^3} \overline{\rho}_a^0 = 1\right\}$ and $F_a^n : \overline{\mathcal{K}}^0_{z_a} \ni \overline{\rho}_a^0 \mapsto -\int_{\mathbb{R}^3} \rho_a^n \log \overline{\rho}_a^0$. Since $\overline{\mathcal{K}}^0_{z_a}$ is a closed finite-dimensional simplex of the vector space $(X^r)^M$, all the relevant topologies on $(X^r)^M$ (the strong and weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, the strong and weak topologies of $L^p(\mathbb{R}^3)$ for $1 < p < \infty$) are equivalent on $\overline{\mathcal{K}}^0_{z_a}$. It is then easily seen using the concavity of the logarithm that the functions F_a^n are convex and continuous, and strictly convex if $\rho_a^n \neq 0$. The same property holds for F_a^*. Let us prove that

$$\overline{\rho}_a^{0,n} \underset{n\to\infty}{\longrightarrow} \overline{\rho}_a^{0,*} \text{ (for any of these topologies)} \quad \Rightarrow \quad F_a^n(\overline{\rho}_a^{0,n}) \underset{n\to\infty}{\longrightarrow} F_a^*(\overline{\rho}_a^{0,*}). \tag{84}$$

First, since for all $n \in \mathbb{N}$, $\overline{\rho}_a^{0,n} := \sum_{k=1}^{m_{z_a}} c_{a,k}^n g_{z_a,k}$ where the $g_{z_a,k}$ are positive L^1-normalized functions of X_+^r and $\sum_{k=1}^{m_{z_a}} c_{a,k}^n = 1$, it holds that $g_a^- \leq \overline{\rho}_a^{0,n} \leq g_a^+$ so that $\log g_a^- \leq \log \overline{\rho}_a^{0,n} \leq \log g_a^+$, where $g_a^- := \min_{1\leq k\leq m_{z_a}} g_{z_a,k}$ and $g_a^+ := \max_{1\leq k\leq m_{z_a}} g_{z_a,k}$. Thus, we obtain that

$$\rho_a^n |\log \overline{\rho}_a^{0,n}| \leq \rho(\cdot + \boldsymbol{R}_a) \max\left(|\log g_a^-|, |\log g_a^+|\right).$$

Assumption (51) implies that $\rho(\cdot + \boldsymbol{R}_a)|\log g_a^-| \in L^1(\mathbb{R}^3)$. In addition, since $g_{z_a,k} \in X_+^r$ for all $1 \leq k \leq m_{z_a}$ and are thus bounded and go to 0 at infinity, it holds that $|\log g_a^+(\boldsymbol{r})| \leq |\log g_a^-(\boldsymbol{r})|$ for sufficiently large values of $|\boldsymbol{r}|$, which implies that $\rho(\cdot + \boldsymbol{R}_a)|\log g_a^+| \in L^1(\mathbb{R}^3)$. Therefore, there exists a nonnegative function $h_a \in L^1(\mathbb{R}^3)$ such that for all $n \in \mathbb{N}$, $\rho_a^n |\log \overline{\rho}_a^{0,n}| \leq h_a$. Let $\epsilon > 0$ and $R > 0$ such that $\int_{\mathbb{R}^3\setminus B_R} h \leq \epsilon$. Then,

$$|F_a^n(\overline{\rho}_a^{0,n}) - F_a^*(\overline{\rho}_a^{0,*})| \leq 2\epsilon + \left|\int_{B_R} \rho_a^n \log \overline{\rho}_a^{0,n} - \int_{B_R} \rho_a^* \log \overline{\rho}_a^{0,*}\right|.$$

Now, denoting by $B_R := \{\mathbf{r} \in \mathbb{R}^3 \,|\, |\mathbf{r}| < R\}$, we have that $(\rho_a^n|_{B_R})_{n\in\mathbb{N}}$ weakly converges to $\rho_a^*|_{B_R}$ in $L^2(B_R)$, while $(\overline{\rho}_a^{0,n})_{n\in\mathbb{N}}$ converges strongly to $\overline{\rho}_a^{0,*}$ since all the relevant topologies are equivalent on $\overline{\mathcal{K}}_{z_a}^0$. The latter property also implies that $(\log \overline{\rho}_a^{0,n})_{n\in\mathbb{N}}$ converges strongly to $\log \overline{\rho}_a^{0,*}$ as we have $g_a^- \leq \overline{\rho}_a^{0,n} \leq g_a^+$ with g_a^+ bounded and g_a^- bounded away from zero on B_R (this function is continuous and positive on the compact $\overline{B}_R$). Hence (84). This properties implies in particular that the sequence of functionals $(F_a^n)_{n\in\mathbb{N}}$ Γ-converges to the functional F_a^*, and therefore that the minimizers of F_a^n converge to the minimizers of F_a^*, these minimizers being unique when ρ_a^n and ρ_a^* are not identically equal to zero. It follows from Lemma 11 that

$$[\boldsymbol{G}^0_{\mathcal{K}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho}^n)]_a = \rho_a^{0,n} = \left| \begin{array}{ll} N_a^n \arg\min_{\overline{\mathcal{K}}_{z_a}^0} F_a^n & \text{if } N_a^n \neq 0, \\ 0 & \text{if } N_a^n = 0, \end{array} \right.$$

converges (in any relevant topologies) to

$$\rho_a^{0,*} = \left| \begin{array}{ll} N_a^* \arg\min_{\overline{\mathcal{K}}_{z_a}^0} F_a^* & \text{if } N_a^* \neq 0, \\ 0 & \text{if } N_a^* = 0, \end{array} \right. = [\boldsymbol{G}^0_{\mathcal{K}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho}^*)]_a.$$

Hence the desired result.

$\boxed{2}$ Let us now prove that $\left(\boldsymbol{\rho}^*, G^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^*)\right)$ is a minimizer to problem (50). To simplify the notation, let us set $\boldsymbol{\rho}^{0,*} = \left(\rho_a^{0,*}\right)_{1\leq a\leq M} := G^0_{\boldsymbol{\mathcal{K}}^0}(\boldsymbol{\rho}^*)$. It is sufficient to prove that for all $1 \leq a \leq M$,

$$s_{\mathrm{KL}}\left(\rho_a^*|\rho_a^{0,*}\right) \leq \liminf_{n\to+\infty} s_{\mathrm{KL}}\left(\rho_a^n|\rho_a^{0,n}\right).$$

(i) Let us again first assume that $N_a^* > 0$. Then, for n large enough, $2N_a^* > N_a^n > N_a^*/2$ and $\overline{\rho}_a^n := \frac{\rho_a^n}{N_a^n}$ defines a probability measure on $\mathbb{R}^3$. Since $\overline{\rho}_a^n \leq 2\frac{\rho}{N_a^*}$ and $\rho \in L^1(\mathbb{R}^3)$, it can be easily checked that the sequence $(\overline{\rho}_a^n)_{n\in\mathbb{N}}$ is tight (in the sense of probability measures). Hence, up to the extraction of a subsequence, $(\overline{\rho}_a^n)_{n\in\mathbb{N}}$ weakly converges in the sense of probability measures to $\overline{\rho}_a^* = \frac{\rho_a^*}{N_a^*}$, i.e. for all continuous bounded functions $f : \mathbb{R}^3 \to \mathbb{R}$,

$$\int_{\mathbb{R}^3} f \, d\overline{\rho}_a^n \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}^3} f \, d\overline{\rho}_a^*.$$

Besides, using similar arguments as above, it holds that the sequence $\left(\overline{\rho}_a^{0,n}\right)_{n\in\mathbb{N}}$ with $\overline{\rho}_a^{0,n} := \frac{\rho_a^{0,n}}{N_a^n}$ weakly converges in the sense of probability measures to $\overline{\rho}_a^{0,*} := \frac{\rho_a^{0,*}}{N_a^*}$. By the joint lower semi-continuity of the Kullback-Leibler divergence for the weak convergence of probability measures, it then holds that

$$\int_{\mathbb{R}^3} \rho_a^* \log(\rho_a^*/\rho_a^{0,*}) \leq \liminf_{n\to+\infty} \int_{\mathbb{R}^3} \rho_a^n \log(\rho_a^n/\rho_a^{0,n}).$$

(ii) Let us now assume that $N_a^* = 0$, which implies that $\rho_a^* = \rho_a^{0,*} = 0$. Since $N_a^n \underset{n\to+\infty}{\longrightarrow} N_a^* = 0$, the sequences $(\rho_a^n)_{n\in\mathbb{N}}$ and $(\rho_a^{0,n})_{n\in\mathbb{N}}$ strongly converge to 0 in $L^1(\mathbb{R}^3)$. Thus, up to the extraction of subsequences, both sequences converge almost everywhere to 0. Using the fact that

- $|x\log x| \leq |y\log y|$ for all $y \leq 1/e$ and $0 \leq x \leq y$;
- $|x\log x| \leq e$ for all $0 \leq x \leq e$,
- $|x\log x| \leq |y\log y|$ for all $y \geq e$ and $0 \leq x \leq y$;

and the bounds $0 \leq \rho_a^n \leq \rho(\cdot + R_a)$ for all $n \in \mathbb{N}$, we obtain that

$$|\rho_a^n \log \rho_a^n| \leq e\chi_{\frac{1}{e}\leq\rho(\cdot+R_a)\leq e} + |\rho(\cdot + R_a)\log\rho(\cdot + R_a)| \left(1 - \chi_{\frac{1}{e}\leq\rho(\cdot+R_a)\leq e}\right).$$

In the case when $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{ISA}}$, it holds that $\rho_a^{0,n} = \langle\rho_a^n\rangle_s$ for all $n \in \mathbb{N}$, so that $0 \leq \rho_a^{0,n} \leq \langle\rho(\cdot + R_a)\rangle_s$. Hence,

$$|\rho_a^{0,n}\log(\rho_a^{0,n})| \leq e\chi_{\frac{1}{e}\leq\langle\rho(\cdot+R_a)\rangle_s\leq e} + |\langle\rho(\cdot + R_a)\rangle_s \log\langle\rho(\cdot + R_a)\rangle_s| \left(1 - \chi_{\frac{1}{e}\leq\langle\rho(\cdot+R_a)\rangle_s\leq e}\right),$$

where χ_E denotes the characteristic function of E. Note also that $\langle\rho(\cdot + R_a)\rangle_s \log\langle\rho(\cdot + R_a)\rangle_s \in L^1(\mathbb{R}^3)$ by the convexity of the function $\mathbb{R}_+ \ni x \mapsto x\log x$. Thus, since

$$\int_{\mathbb{R}^3} \rho_a^n \log\left(\frac{\rho_a^n}{\langle\rho_a^n\rangle_s}\right) = \int_{\mathbb{R}^3} \rho_a^n \log(\rho_a^n) - \int_{\mathbb{R}^3} \langle\rho_a^n\rangle_s \log(\langle\rho_a^n\rangle_s),$$

we deduce from the Lebesgue dominated convergence theorem that

$$\int_{\mathbb{R}^3} \rho_a^n \log\left(\frac{\rho_a^n}{\langle\rho_a^n\rangle_s}\right) \underset{n\to+\infty}{\longrightarrow} 0 = \int_{\mathbb{R}^3} \rho_a^* \log\left(\frac{\rho_a^*}{\langle\rho_a^*\rangle_s}\right).$$

In the case when $\boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}$, since for all $n \in \mathbb{N}$, $\rho_a^{0,n} := \sum_{k=1}^{m_{z_a}} c_{a,k}^n g_{z_a,k}$ where the $g_{z_a,k}$ are positive L^1-normalized functions of X_+^r, it holds that $N_a^n g_a^- \leq \rho_a^{0,n} \leq N_a^n g_a^+$ so that $\log g_a^- + \log N_a^n \leq \log\rho_a^{0,n} \leq \log g_a^+ + \log N_a^n$. As a consequence,

$$|\log\rho_a^{0,n}| \leq |\log N_a^n| + \max\left(|\log g_a^-|, |\log g_a^+|\right),$$

and

$$\rho_a^n|\log\rho_a^{0,n}| \leq \rho_a^n|\log N_a^n| + \rho_a^n \max\left(|\log g_a^-|, |\log g_a^+|\right).$$

On the one hand,

$$\int_{\mathbb{R}^3} \rho_a^n|\log N_a^n| = N_a^n|\log N_a^n| \underset{n\to+\infty}{\longrightarrow} 0.$$

On the other hand, $\rho_a^n \max\left(|\log g_a^-|, |\log g_a^+|\right) \leq \rho(\cdot + \mathbf{R}_a)\max\left(|\log g_a^-|, |\log g_a^+|\right) \in L^1(\mathbb{R}^3)$. Thus, the Lebesgue dominated convergence theorem yields that

$$\int_{\mathbb{R}^3} \rho_a^n \max\left(|\log g_a^-|, |\log g_a^+|\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Finally, $\int_{\mathbb{R}^3} \rho_a^n |\log \rho_a^{0,n}| \longrightarrow_{n\to+\infty} 0$, and

$$\int_{\mathbb{R}^3} \rho_a^n \log\left(\frac{\rho_a^n}{\rho_a^{0,n}}\right) \underset{n\to+\infty}{\longrightarrow} 0 = \int_{\mathbb{R}^3} \rho_a^* \log\left(\frac{\rho_a^*}{\rho_a^{0,*}}\right).$$

We have thus proved that in any case,

$$\int_{\mathbb{R}^3} \rho_a^* \log\left(\frac{\rho_a^*}{\langle \rho_a^* \rangle_s}\right) \leq \liminf_{n\to+\infty} \int_{\mathbb{R}^3} \rho_a^n \log\left(\frac{\rho_a^n}{\langle \rho_a^n \rangle_s}\right)$$

for all $1 \leq a \leq M$, which implies that $\left(\boldsymbol{\rho}^*, \boldsymbol{\rho}^{0,*}\right)$ is a minimizer to (50).

$\boxed{3}$ The uniqueness comes from the strict convexity of the functional

$$\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0 \ni \left(\boldsymbol{\rho}, \boldsymbol{\rho}^0\right) \mapsto S(\boldsymbol{\rho}|\boldsymbol{\rho}^0) \in \mathbb{R}_+ \cup \{+\infty\}$$

on the convex subset of $\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}} \times \boldsymbol{\mathcal{K}}^0$ on which it is finite.

Lastly, (52) is a direct consequence of the fact that

$$\boldsymbol{\rho}^* := \operatorname{argmin}_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S\left(\boldsymbol{\rho}|\boldsymbol{\rho}^{0,*}\right)$$

and Lemma 2.

6.6 Proof of Theorem 7

It is easy to check that the assumptions $\rho > 0$ and $s_{\mathrm{KL}}(\rho|\rho^{0,(0)}) < +\infty$ ensure that for both ISA and L-ISA, the first iteration is well defined. Assume now by induction that the first $\widetilde{m}$ iterations are well-defined for some $\widetilde{m} \in \mathbb{N}^*$ and that $s_{\mathrm{KL}}(\rho|\rho^{0,(m-1)}) < +\infty$ for all $1 \leq m \leq \widetilde{m}$.

For all $1 \leq m \leq \widetilde{m}$ and $1 \leq a \leq M$, we denote by

$$N_a^{(m)} := \int_{\mathbb{R}^3} \rho_a^{(m)} = \int_{\mathbb{R}^3} \rho_a^{0,(m)} \quad \text{and} \quad \mathbf{N}^{(m)} = (N_a^{(m)})_{1\leq a\leq M} \in \mathbb{R}_+^M.$$

From the assumptions of Theorem 7, it can be easily proven by recursion that $N_a^{(m)} > 0$ for all $1 \leq m \leq \widetilde{m}$ and all $1 \leq a \leq M$.

Using Lemmata 2 and 11, we obtain that for all $1 \leq m \leq \widetilde{m}$, $\boldsymbol{\rho}^{(m)}$ and $\boldsymbol{\rho}^{0,(m)}$ are respectively minimizers of

$$\boldsymbol{\rho}^{(m)} \in \operatorname*{argmin}_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^{0,(m-1)}) \tag{85}$$

and

$$\boldsymbol{\rho}^{0,(m)} \in \operatorname*{argmin}_{\boldsymbol{\rho}^0\in\boldsymbol{\mathcal{K}}^0,\, \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)=\mathbf{N}^{(m)}} S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^0). \tag{86}$$

In order to prove (54), we first show that

$$S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) \geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)}). \tag{87}$$

It follows from Lemmata 2 and 11 that $\boldsymbol{\mu}^{(m)} \odot \boldsymbol{\rho}^{0,(m)}$, with $\boldsymbol{\mu}^{(m)} := \left(\frac{N_a^{(m-1)}}{N_a^{(m)}}\right)_{1\leq a\leq M}$, is the unique minimizer of

$$\inf_{\boldsymbol{\rho}^0\in\boldsymbol{\mathcal{K}}^0,\, \boldsymbol{\mathcal{N}}(\boldsymbol{\rho}^0)=\mathbf{N}^{(m-1)}} S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^0). \tag{88}$$

We therefore have

$$\begin{aligned}
S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) &\geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\mu}^{(m)} \odot \boldsymbol{\rho}^{0,(m)}) \\
&= \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a^{(m)} \log\left(\frac{\rho_a^{(m)}}{\frac{N_a^{(m-1)}}{N_a^{(m)}}\rho_a^{0,(m)}}\right) \\
&= \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a^{(m)} \log\left(\frac{\rho_a^{(m)}}{\rho_a^{0,(m)}}\right) + \int_{\mathbb{R}^3} \rho_a^{(m)} \log\left(\frac{N_a^{(m)}}{N_a^{(m-1)}}\right) \\
&= \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a^{(m)} \log\left(\frac{\rho_a^{(m)}}{\rho_a^{0,(m)}}\right) + N_a^{(m)} \log\left(\frac{N_a^{(m)}}{N_a^{(m-1)}}\right), \\
&= S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)}) + \sum_{a=1}^{M} N_a^{(m)} \log\left(\frac{N_a^{(m)}}{N_a^{(m-1)}}\right).
\end{aligned}$$

It thus remains to show that

$$\sum_{a=1}^{M} N_a^{(m)} \log\left(\frac{N_a^{(m)}}{N_a^{(m-1)}}\right) \geq 0.$$

We now use the fact that $-\log x \geq 1 - x$ (since $\log x \leq x - 1$) to get

$$\begin{aligned}
\sum_{a=1}^{M} N_a^{(m)} \log\left(\frac{N_a^{(m)}}{N_a^{(m-1)}}\right) &= \sum_{a=1}^{M} -N_a^{(m)} \log\left(\frac{N_a^{(m-1)}}{N_a^{(m)}}\right) \\
&\geq \sum_{a=1}^{M} N_a^{(m)} \left(1 - \frac{N_a^{(m-1)}}{N_a^{(m)}}\right) = \sum_{a=1}^{M} (N_a^{(m)} - N_a^{(m-1)}) = 0,
\end{aligned}$$

since

$$\sum_{a=1}^{M} N_a^{(m)} = \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a^{(m)} = \int_{\mathbb{R}^3} \rho = \sum_{a=1}^{M} \int_{\mathbb{R}^3} \rho_a^{(m-1)} = \sum_{a=1}^{M} N_a^{(m-1)}.$$

Hence (87). Let us now show that

$$S(\boldsymbol{\rho}^{(m-1)}|\boldsymbol{\rho}^{0,(m-1)}) \geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) + \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^{M} \left\|\rho_a^{(m)} - \rho_a^{(m-1)}\right\|_{L^2}^2. \tag{89}$$

Note that, by definition of $\boldsymbol{\rho}^{(m)}$, $S(\boldsymbol{\rho}^{(m-1)}|\boldsymbol{\rho}^{0,(m-1)}) \geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)})$. Using a second-order Taylor expansion formula with integral remainder, it holds that for all $x, x', y > 0$,

$$x' \log\left(\frac{x'}{y}\right) \geq x \log\left(\frac{x}{y}\right) + (x' - x)\left[\log\left(\frac{x}{y}\right) + 1\right] + \frac{1}{\max(x, x')} \frac{(x - x')^2}{2}.$$

As a consequence, since for all $1 \leq a \leq M$, $\frac{1}{\|\rho_a^{(m)}\|_{L^\infty}} \geq \frac{1}{\|\rho\|_{L^\infty}}$, we obtain that

$$\begin{aligned}
S(\boldsymbol{\rho}^{(m-1)}|\boldsymbol{\rho}^{0,(m-1)}) \geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) &+ \sum_{a=1}^{M} \int_{\mathbb{R}^3} (\rho_a^{(m-1)} - \rho_a^{(m)}) \left[\log\left(\frac{\rho_a^{(m)}}{\rho_a^{0,(m-1)}}\right) + 1\right] \\
&+ \sum_{a=1}^{M} \frac{1}{2\|\rho\|_{L^\infty}} \int_{\mathbb{R}^3} (\rho_a^{(m-1)} - \rho_a^{(m)})^2.
\end{aligned}$$

Using the fact that the expression

$$\frac{\rho_a^{(m)}(\boldsymbol{r})}{\rho_a^{0,(m-1)}(\boldsymbol{r})} = \frac{\rho(\boldsymbol{r}+\mathbf{R}_a)}{\sum_{b=1}^M \rho_b^{0,(m-1)}(\boldsymbol{r}-\mathbf{R}_b+\mathbf{R}_a)}$$

is independent of a and the equality

$$\sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a^{(m-1)} = \sum_{a=1}^M \int_{\mathbb{R}^3} \rho_a^{(m)} = \int_{\mathbb{R}^3} \rho,$$

we thus obtain that

$$\begin{aligned}
S(\boldsymbol{\rho}^{(m-1)}|\boldsymbol{\rho}^{0,(m-1)}) &\geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) \\
&+ \sum_{a=1}^M \int_{\mathbb{R}^3} (\rho_a^{(m-1)}(\boldsymbol{r}-\mathbf{R}_a) - \rho_a^{(m)}(\boldsymbol{r}-\mathbf{R}_a)) \log\left(\frac{\rho(\boldsymbol{r})}{\sum_{b=1}^M \rho_b^{0,(m-1)}(\boldsymbol{r}-\mathbf{R}_b)}\right) d\boldsymbol{r} \\
&+ \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^M \left\|\rho_a^{(m-1)} - \rho_a^{(m)}\right\|_{L^2}^2, \\
&= S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) + \int_{\mathbb{R}^3} (\rho(\boldsymbol{r}) - \rho(\boldsymbol{r})) \log\left(\frac{\rho(\boldsymbol{r})}{\sum_{b=1}^M \rho_b^{0,(m-1)}(\boldsymbol{r}-\mathbf{R}_b)}\right) d\boldsymbol{r} \\
&+ \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^M \left\|\rho_a^{(m-1)} - \rho_a^{(m)}\right\|_{L^2}^2, \\
&= S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) + \frac{1}{2\|\rho\|_{L^\infty}} \sum_{a=1}^M \left\|\rho_a^{(m-1)} - \rho_a^{(m)}\right\|_{L^2}^2.
\end{aligned}$$

Hence (89). Collecting this result together with (87), we obtain (54). It also follows from (17), (85), and (87) that

$$s_{\mathrm{KL}}(\rho, \rho^{0,(m-1)}) = S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m-1)}) \geq S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)}) \geq \inf_{\boldsymbol{\rho}\in\boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}} S(\boldsymbol{\rho}|\boldsymbol{\rho}^{0,(m)}) = s_{\mathrm{KL}}(\rho, \rho^{0,(m)}).$$

In particular $s_{\mathrm{KL}}(\rho, \rho^{0,(m)}) < +\infty$ for all $1 \leq m \leq \widetilde{m}$, so that the iteration $\widetilde{m}+1$ is well defined. By recursion, all the iterations are well defined and the above results are valid for all $m \geq 1$. The bound (55) is obtained by a simple combination of the above inequalities.

Another consequence of (54) is that $\left(S(\boldsymbol{\rho}^{(m)}|\boldsymbol{\rho}^{0,(m)})\right)_{m\geq 0}$ is a non-increasing sequence of nonnegative real numbers, hence converges to a nonnegative real number as m tends to infinity. The bound (54) also implies that

$$\sum_{m\geq 1} \sum_{a=1}^M \left\|\rho_a^{(m)} - \rho_a^{(m-1)}\right\|_{L^2}^2 < +\infty,$$

which, in turn, implies (56).

Let us finally prove the last assertion of Theorem 7. We infer from Lemma 12 that there exist $\boldsymbol{\rho}^* \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$ and a subsequence $(\boldsymbol{\rho}^{(m_l)})_{l\in\mathbb{N}}$ of $(\boldsymbol{\rho}^{(m)})_{m\in\mathbb{N}}$ converging to $\boldsymbol{\rho}^*$ for the weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, and for the weak topology of $L^p(\mathbb{R}^3)^M$ for any $1 < p < \infty$, and such that

$$\forall 1 \leq a \leq M, \quad N_a^{(m_l)} := \int_{\mathbb{R}^3} \rho_a^{(m_l)}(\mathbf{r})\, dr \underset{l\to\infty}{\longrightarrow} \int_{\mathbb{R}^3} \rho_a^*(\mathbf{r})\, dr =: N_a^*.$$

In addition, we can prove that there exists $\boldsymbol{\rho}^{0,*} \in \boldsymbol{\mathcal{K}}^0$, such that, up to extraction of a subsequence, $(\boldsymbol{\rho}^{0,(m_l)})_{l\in\mathbb{N}}$ converges to $\boldsymbol{\rho}^{0,*}$ for the weak-* topologies of $\mathcal{M}_{\mathrm{b}}(\mathbb{R}^3)^M$ and $L^\infty(\mathbb{R}^3)^M$, and for the weak topology of $L^p(\mathbb{R}^3)^M$ for any $1 < p < \infty$. Using similar arguments as in the proof of Theorem 6, we obtain that

$$\boldsymbol{\rho}^{0,*} = \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho}^*).$$

Moreover, since $\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}} \subset \mathrm{Span}\,\{g_{z_a,k}, 1 \le a \le M,\ 1 \le k \le m_{z_a}\}$, which is a finite-dimensional vector space, the convergence of the sequence $(\boldsymbol{\rho}^{0,(m_l)})_{l\in\mathbb{N}}$ to $\boldsymbol{\rho}^{0,*}$ holds strongly in any $L^p(\mathbb{R}^3)$ for any $1 \le p \le +\infty$.

Using the fact that

$$\rho_a^{(m_l+1)}(\boldsymbol{r}) = \frac{\rho_a^{0,(m_l)}(\boldsymbol{r})}{\sum_{b=1} \rho_b^{0,(m_l)}(\boldsymbol{r} - \mathbf{R}_b + \mathbf{R}_a)} \rho(\boldsymbol{r} + \mathbf{R}_a)$$

together with (56) yields that

$$\rho_a^*(\boldsymbol{r}) = \frac{\rho_a^{0,*}(\boldsymbol{r})}{\sum_{b=1} \rho_b^{0,*}(\boldsymbol{r} - \mathbf{R}_b + \mathbf{R}_a)} \rho(\boldsymbol{r} + \mathbf{R}_a).$$

This relationship, together with the fact that $\boldsymbol{\rho}^{0,*} = \boldsymbol{G}^0_{\boldsymbol{\mathcal{K}}^0_{\mathrm{L-ISA}}}(\boldsymbol{\rho}^*)$ implies that $(\boldsymbol{\rho}^*, \boldsymbol{\rho}^{0,*})$ is the unique minimizer of (50). Hence the desired result.

6.7 Proof of Proposition 8

If $\rho(\cdot + \mathbf{R}_2) = \langle \rho(\cdot + \mathbf{R}_2)\rangle_s$, it is clear that $\rho_1^{\mathrm{opt}} = 0$ and $\rho_2^{\mathrm{opt}} = \langle \rho^{\mathrm{opt}}\rangle_s$.

Conversely, let us assume that $\rho_1^{\mathrm{opt}} = 0$ (and therefore $\rho_2^{\mathrm{opt}} = \rho(\cdot + \mathbf{R}_2) > 0$ on $\mathbb{R}^3$, which implies that $\langle \rho_2^{\mathrm{opt}}\rangle_s = \langle \rho(\cdot + \mathbf{R}_2)\rangle_s > 0$ on $\mathbb{R}^3$), and prove that necessarily $\rho_2^{\mathrm{opt}} = \langle \rho(\cdot + \mathbf{R}_2)\rangle_s$. For all $\boldsymbol{\rho} \in \boldsymbol{\mathcal{K}}_{\rho,\mathbf{R}}$, we denote by $\mathcal{J}(\boldsymbol{\rho}) := S(\boldsymbol{\rho}|\langle\boldsymbol{\rho}\rangle_s)$. Let us consider radially symmetric functions $h_1 \in X^r$ and $h_2 \in X^r$ such that $\langle \rho_2^{\mathrm{opt}}\rangle_s + \epsilon h_2 > 0$ a.e. for $\epsilon > 0$ small enough, and consider the perturbed AIM densities

$$\rho_a^\epsilon := \frac{\langle \rho_a^{\mathrm{opt}}\rangle_s + \epsilon h_a}{\sum_b (\langle \rho_b^{\mathrm{opt}}\rangle_s + \epsilon h_b)(\cdot - \mathbf{R}_b + \mathbf{R}_a)} \rho(\cdot + \mathbf{R}_a).$$

To the first order in ϵ, we have, uniformly in $L^\infty_{\mathrm{loc}}(\mathbb{R}^3)$,

$$\begin{aligned}
\rho_1^\epsilon &= \epsilon \frac{h_1}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s(\cdot + \mathbf{R}_1)} \rho(\cdot + \mathbf{R}_1) + o(\epsilon),\\
\rho_2^\epsilon &= \rho(\cdot + \mathbf{R}_2) - \epsilon \frac{h_1(\cdot - \mathbf{R}_1 + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \rho(\cdot + \mathbf{R}_2) + o(\epsilon),
\end{aligned}$$

and thus

$$\mathcal{J}(\boldsymbol{\rho}^\epsilon) - \mathcal{J}(\boldsymbol{\rho}^{\mathrm{opt}}) = -\epsilon \int_{\mathbb{R}^3} \frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right) h_1(\cdot - \mathbf{R}_1 + \mathbf{R}_2) + o(\epsilon).$$

Let us denote by

$$\rho^{s,2}(\boldsymbol{r}) := \langle \rho(\cdot + \mathbf{R}_2)\rangle_s (\boldsymbol{r} - \mathbf{R}_2).$$

It then holds that

$$\begin{aligned}\mathcal{J}(\boldsymbol{\rho}^\epsilon) - \mathcal{J}(\boldsymbol{\rho}^{\mathrm{opt}}) &= -\epsilon \int_{\mathbb{R}^3} \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right) h_1 + o(\epsilon) \\ &= -\epsilon \int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s h_1 + o(\epsilon)\end{aligned}$$

First, we have

$$\begin{aligned}\int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s &= \int_{\mathbb{R}^3} \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right) \\ &= \int_{\mathbb{R}^3} \frac{\rho}{\rho^{s,2}} \log\left(\frac{\rho}{\rho^{s,2}}\right) \\ &= \int_{\mathbb{R}^3} \left(\frac{\rho}{\rho^{s,2}} \log\left(\frac{\rho}{\rho^{s,2}}\right)\right)^{s,2} \\ &= \int_{\mathbb{R}^3} \frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right) \\ &= \int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right)\right\rangle_s \\ &\geq \int_{\mathbb{R}^3} \frac{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right) = 0,\end{aligned}$$

the last inequality being a consequence of Jensen's inequality. Let us reason by contradiction and assume that

$$\int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s > 0.$$

There would then exist a bounded Borel subset A of $[0, +\infty)$ with positive Lebesgue measure such that $w(r) > 0$ for almost all $r \in A$ where $w \in Y_+$ is the function such that

$$\forall r > 0,\ \forall \boldsymbol{\sigma} \in \mathbb{S}^2, \quad w(r) := \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s (r\boldsymbol{\sigma}).$$

Choosing a particular function h_1 such that $h_1(r\boldsymbol{\sigma}) = g_1(r)$ with $g_1 > 0$ on A and $h_1 = 0$ on $[0, +\infty) \setminus A$ would imply that

$$\int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s h_1 = 4\pi \int_0^{+\infty} w(r) g_1(r) r^2 \, dr > 0.$$

Letting ϵ go to 0, we reach a contradiction since $(\boldsymbol{\rho}^{\mathrm{opt}}, \langle \boldsymbol{\rho}^{\mathrm{opt}}\rangle_s)$ is a minimizer of (50).

Hence, it necessarily holds that

$$\int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)} \log\left(\frac{\rho(\cdot + \mathbf{R}_1)}{\rho^{s,2}(\cdot + \mathbf{R}_1)}\right)\right\rangle_s = \int_{\mathbb{R}^3} \left\langle \frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right)\right\rangle_s = 0.$$

Since $\left\langle \frac{\rho(\cdot+\mathbf{R}_2)}{\langle \rho(\cdot+\mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot+\mathbf{R}_2)}{\langle \rho(\cdot+\mathbf{R}_2)\rangle_s}\right)\right\rangle_s \geq 0$ almost everywhere from Jensen's inequality, we obtain

$$\left\langle \frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s} \log\left(\frac{\rho(\cdot + \mathbf{R}_2)}{\langle \rho(\cdot + \mathbf{R}_2)\rangle_s}\right)\right\rangle_s = 0.$$

This equality implies that

$$\langle \rho(\cdot + \mathbf{R}_2) \log \rho(\cdot + \mathbf{R}_2)\rangle_s = \langle \rho(\cdot + \mathbf{R}_2)\rangle_s \log\langle \rho(\cdot + \mathbf{R}_2)\rangle_s \text{ a.e.}$$

Using again Jensen's inequality and the strict convexity of the function $x \mapsto x \log x$, we finally obtain $\rho(\cdot + \mathbf{R}_2) = \langle \rho(\cdot + \mathbf{R}_2)\rangle_s$. Hence the desired result.

6.8 Proof of Proposition 9

It is clear from their definition that the functions w_a are nonnegative and bounded by M_ρ. Taking spherical averages in the optimality condition (52), and using (53), we see that for all $1 \le a \le 2$,

$$I_a(r) = \fint_{\mathbb{S}^2} \frac{\rho(\mathbf{R}_a + r\boldsymbol{\sigma})}{w_a(r) + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|)}\, d\boldsymbol{\sigma} = 1, \quad \text{or} \quad w_a(r) = 0,$$

where $b = 2$ if $a = 1$ and $b = 1$ if $a = 2$. For given and fixed w_b, the function w_a is therefore defined implicitly on $(0, +\infty)$ by

$$\text{for almost all } r \in (0, +\infty), \quad F_a(r, w_a(r)) = 0 \quad \text{or} \quad w_a(r) = 0,$$

where the function $F_a : (0, +\infty) \times \mathbb{R}_+ \to [0, +\infty]$ is given by

$$F_a(r, y) = \fint_{\mathbb{S}^2} \frac{\rho(\mathbf{R}_a + r\boldsymbol{\sigma})}{y + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|)}\, d\boldsymbol{\sigma} - 1.$$

Note that the function $\mathbb{S}^2 \ni \boldsymbol{\sigma} \mapsto \rho(\mathbf{R}_a + r\boldsymbol{\sigma})$ is continuous on $\mathbb{S}^2$, nonnegative, and bounded by M_ρ. Consequently, F_a is well-defined on $(0, +\infty) \times (0, +\infty)$, and for each $r > 0$, the function $(0, +\infty) \ni y \mapsto F_a(r, y) \in \mathbb{R}$ is decreasing and converges to -1 at infinity, so that for each $r \in (0, +\infty)$, there exists at most one $y \in \mathbb{R}_+$ such that $F_a(r, y) = 0$. We denote by $w_a^+(r)$ the unique solution in $\mathbb{R}_+$ to $F_a(r, w_a^+(r)) = 0$ if it exists, and we set $w_a^+(r) = 0$ otherwise. We therefore have $w_a(r) \in \{0, w_a^+(r)\}$ for almost all $r \in \mathbb{R}_+$.

We have for all $r, y, z \in (0, +\infty)$,

$$\begin{aligned} |F_a(r, y) - F_a(r, z)| &= \left| \fint_{\mathbb{S}^2} \rho(\mathbf{R}_a + r\boldsymbol{\sigma}) \left(\frac{1}{y + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|)} - \frac{1}{z + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|)} \right) d\boldsymbol{\sigma} \right| \\ &= K_a(r, y, z)\, |y - z|, \end{aligned} \tag{90}$$

with

$$K_a(r, y, z) := \fint_{\mathbb{S}^2} \frac{\rho(\mathbf{R}_a + r\boldsymbol{\sigma})}{(y + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|))(z + w_b(|\mathbf{R}_a + r\boldsymbol{\sigma} - \mathbf{R}_b|))}\, d\boldsymbol{\sigma}.$$

Using the fact that ρ is continuous and positive everywhere on $\mathbb{R}^3$, hence bounded away from zero on any compact subset of $\mathbb{R}^3$, we get that for all $(r, y, z) \in (0, +\infty) \times [0, M_\rho]^2$,

$$K_a(r, y, z) \ge \frac{m_a(r)}{2M_\rho^2} > 0$$

where

$$m_a(r) := \min_{\overline{B}(0, r + |\mathbf{R}_a|)} \rho$$

is a continuous, positive, decreasing function of r.

For $r_1, r_2 \in \mathbb{R}_+$ such that $r_1 + r_2 \ge R = |\mathbf{R}_a - \mathbf{R}_b|$, we denote by $\theta_1(r_1, r_2)$ and $\theta_2(r_1, r_2)$, the unique real numbers in $[0, \pi]$ such that

$$r_1^2 + R^2 - 2r_1 R \cos(\theta_1(r_1, r_2)) = r_2^2 \quad \text{and} \quad r_2^2 + R^2 - 2r_2 R \cos(\theta_2(r_1, r_2)) = r_1^2,$$

and by $\widetilde{\rho}_a(r, \theta, \phi)$ the representation of the total density ρ in spherical coordinates relative to $\mathbf{R}_a$, with the z-axis aligned with $\mathbf{e} := \frac{\mathbf{R}_b - \mathbf{R}_a}{R}$, i.e.

$$\widetilde{\rho}_a(r, \theta, \phi) := \rho(\mathbf{R}_a + r(\sin\theta\cos\phi\, \mathbf{e}_x + \sin\theta\sin\phi\, \mathbf{e}_y + \cos\theta\, \mathbf{e})),$$

where $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e})$ forms an orthonormal basis of $\mathbb{R}^3$. Using the change of variables $r_2 = (r_1^2 + R^2 - 2r_1 R\cos\theta)^{1/2}$, for which $r_1 R \sin\theta\, d\theta = r_2\, dr_2$, we have

$$\begin{aligned} I_1(r_1) &= \frac{1}{4\pi}\int_0^{\pi} d\theta \int_0^{2\pi} d\phi\, \frac{\widetilde{\rho}_1(r_1,\theta,\phi)}{w_1(r_1) + w_2((r_1^2 + R^2 - 2r_1 R\cos\theta)^{1/2})}\, \sin\theta \\ &= \frac{1}{4\pi}\int_{|R-r_1|}^{R+r_1} dr_2 \int_0^{2\pi} d\phi\, \frac{\widetilde{\rho}_1(r_1,\theta_1(r_1,r_2),\phi)}{w_1(r_1) + w_2(r_2)}\, \frac{r_2}{r_1 R} \\ &= \int_{|R-r_1|}^{R+r_1} \frac{h_1(r_1,r_2)}{w_1(r_1) + w_2(r_2)}\, dr_2, \end{aligned}$$

where the function $h_1 : D_1 \to \mathbb{R}_+$ is defined by

$$D_1 := \{(r_1,r_2),\ r_1 \in (0,+\infty),\ r_2 \in [|R-r_1|, R+r_1]\}, \quad h_1(r_1,r_2) := \frac{r_2}{4\pi r_1 R}\int_0^{2\pi} \widetilde{\rho}_1(r_1,\theta_1(r_1,r_2),\phi)\, d\phi.$$

The function w_1 is therefore defined implicitly on $(0,+\infty)$ by

$$F_1(r_1, w_1(r_1)) = 0 \quad \text{or} \quad w_1(r_1) = 0,$$

where the function $F_1 : (0,+\infty) \times \mathbb{R}_+ \to [0,+\infty]$ is given by

$$F_1(r_1, y) = \int_{|R-r_1|}^{R+r_1} \frac{h_1(r_1,r_2)}{y + w_2(r_2)}\, dr_2 - 1.$$

Note that the function h_1 is continuous on D_1 and that we have for all $(r_1,r_2) \in D_1$,

$$|h_1(r_1,r_2)| \le \frac{M_\rho r_2}{2r_1 R}.$$

By dominated convergence, F_1 is therefore locally bounded and continuous on $(0,+\infty)\times(0,+\infty)$.

Recall that $0 \le w_a(r_a) \le M_\rho$ for a.a. $r_a \in (0,+\infty)$. We therefore have for all $r_1, r_1', y, z \in (0,+\infty)$,

$$\begin{aligned} \big|F_1(r_1,y) - F_1(r_1',z)\big| &\le \big|F_1(r_1,y) - F_1(r_1',y)\big| + \big|F_1(r_1',y) - F_1(r_1',z)\big| \\ &\le \frac{2M_\rho}{\min(y,z)}|r_1 - r_1'| + \frac{M_\rho}{yz}|y-z|, \end{aligned}$$

which proves that F_1 is locally Lipschitz on the open set $(0,+\infty)\times(0,+\infty)$. It follows from (90) and the implicit function theorem for Lipschitz functions (see Theorem 13 in Appendix B) that if $w_1^+(r_{1*}) > 0$ for some $r_{1*} > 0$, then w_1^+ is Lipschitz in the neighborhood of r_{1*}. Let $r_{1*} \in (0,+\infty)$ be such that $w_1^+(r_{1*}) = 0$, and $(r_{1,n}) \in (0,+\infty)^{\mathbb{N}}$ a sequence converging to r_{1*}. We have for all $n \in \mathbb{N}$

$$F_1(r_{1,n}, w_1^+(r_{1,n})) = \int_{|R-r_{1,n}|}^{R+r_{1,n}} \frac{h_1(r_{1,n},r_2)}{w_1^+(r_{1,n}) + w_2(r_2)}\, dr_2 - 1 = 0.$$

Assume that we can extract from $(r_{1,n})_{n\in\mathbb{N}}$ a subsequence $(r_{1,n_k})_{k\in\mathbb{N}}$ with positive limit, i.e., such that $w_1^+(r_{1,n_k}) \to y_* > 0$ as k tends to infinity. Passing to the limit by the dominated convergence theorem, we get

$$F_1(r_{1*}, y_*) = \int_{|R-r_{1*}|}^{R+r_{1*}} \frac{h_1(r_{1*},r_2)}{y_* + w_2(r_2)}\, dr_2 - 1 = 0.$$

Hence, $y_* = w_1^+(r_{1*}) = 0$. We reach a contradiction. This proves that w_1^+ is continuous on $(0, +\infty)$. The same holds for w_2^+ by symmetry. This completes the proof of the first statement.

Let us now establish the second one. If the w_a's are bounded away from zero on every compact subset of $[0, +\infty)$, then $w_a = w_a^+$ almost everywhere on $[0, +\infty)$, which, in view of the first statement, implies that the w_a's are Lipschitz on $(0, +\infty)$ and that for all $r_1, r_2 > 0$,

$$F_1(r_1, w_1(r_1)) = 0 \quad \text{and} \quad F_2(r_2, w_2(r_2)) = 0$$

where

$$F_1(r_1, y) = \frac{1}{4\pi} \int_{\mathbb{S}^2} \frac{\rho(\mathbf{R}_1 + r_1\boldsymbol{\sigma})}{y + w_2(|r_1\boldsymbol{\sigma} - R\mathbf{e}|)} \, d\boldsymbol{\sigma} - 1,$$
$$F_2(r_2, y) = \frac{1}{4\pi} \int_{\mathbb{S}^2} \frac{\rho(\mathbf{R}_2 + r_2\boldsymbol{\sigma})}{w_1(|r_2\boldsymbol{\sigma} + R\mathbf{e}|) + y} \, d\boldsymbol{\sigma} - 1.$$

By a simple continuity argument, we deduce from the fact that the w_j's are bounded away from zero on $[0, 1]$ that w_j is Lipschitz on the whole interval $[0, +\infty)$. If ρ is C^1 away from the centers $\mathbf{R}_j$, then the functions $F_j : [0, +\infty) \times (0, +\infty) \to \mathbb{R}$ are C^1 on $((0, R) \cup (R, +\infty)) \times (0, +\infty)$. By the implicit function theorem for C^1 functions, the functions w_j are C^1 on $(0, R) \cup (R, +\infty)$ (we use (90) and its analogue for w_2 to show that the assumptions of the implicit function theorem are satisfied). By a simple bootstrap argument, if ρ is C^k away from the centers $\mathbf{R}_j$, then the functions w_j are C^k on $(0, R) \cup (R, +\infty)$.

Ackowledgements

R.B. thanks Michael Herbst, Emmanuel Giner and Laurent Vidal for useful discussions and technical help. The authors are also very grateful to Antoine Levitt and Alston Misquitta for stimulating discussions. This publication is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement $n°$ 810367).

Appendix A: Non-uniqueness of GISA fixed-points

Since GISA is covered by the framework (50), uniqueness of a minimizer, as stated in Theorem 6 for ISA and L-ISA, can not be guaranteed. We have therefore investigated the uniqueness of minimizers and have observed numerically a sensitivity of the GISA solution (obtained by the fixed-point iterations described in Section 2.6) to the initial guess that we report in this Appendix. More specifically, for a (non-dimensional) test density sum of two normalized Gaussian functions:

$$\rho(\mathbf{r}) = \left(\frac{\alpha_1}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_1|\mathbf{r}-\mathbf{R_1}|^2} + \left(\frac{\alpha_2}{\pi}\right)^{\frac{3}{2}} e^{-\alpha_2|\mathbf{r}-\mathbf{R_2}|^2} \tag{91}$$

with $\alpha_1 = 0.1$, $\alpha_2 = 0.5$, $|\mathbf{R}_1 - \mathbf{R}_2| = 1.131$, using six shells on each atom in the GISA pro-atomic densities variational space, and Gaussian exponents

$$\alpha_{1,k} \in \{0.01, 0.1, 1, 2, 5, 10\}, \quad \alpha_{2,k} \in \{0.05, 0.5, 2, 4, 10, 50\},$$

we found two local minima $\rho_a^{0,*}$ and $\rho_a^{0,**}$, whose profiles $r \mapsto \log\left(4\pi r^2 \rho_a^0(r)\right)$ are represented in Figure 7. The expected solution, corresponding to one unit charge on each atom was obtained e.g. with the initial guess:

$$\left(c_{1,k}^{(0),*}\right)_{k=1..6} = (0, 0, 0, 1, 0, 0), \quad \left(c_{2,k}^{(0),*}\right)_{k=1..6} = (0, 0, 0, 0, 1, 0), \tag{92}$$

(*i.e.* with weights initially on more compact Gaussians than the expected solutions), while a more balanced initial guess:

$$\left(c_{1,k}^{(0),**}\right)_{k=1..6} = \left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right), \quad \left(c_{2,k}^{(0),**}\right)_{k=1..6} = \left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right) \tag{93}$$

leads to a fixed-point associated to slightly different pro-atomic densities (see Figure 7, right panel) and slightly distorted atomic charges and dipoles, see Table 3.

	Fixed point 1 (expected solution)	Fixed point 2
Charge q_1	1.000	0.977
Charge q_2	1.000	1.023
Dipole d_z^1	0.000	0.020
Dipole d_z^2	0.000	0.006

Table 3: Local charges and dipoles (component along the z axis) on the two atoms computed by GISA from two different initial guesses (92) and (93).

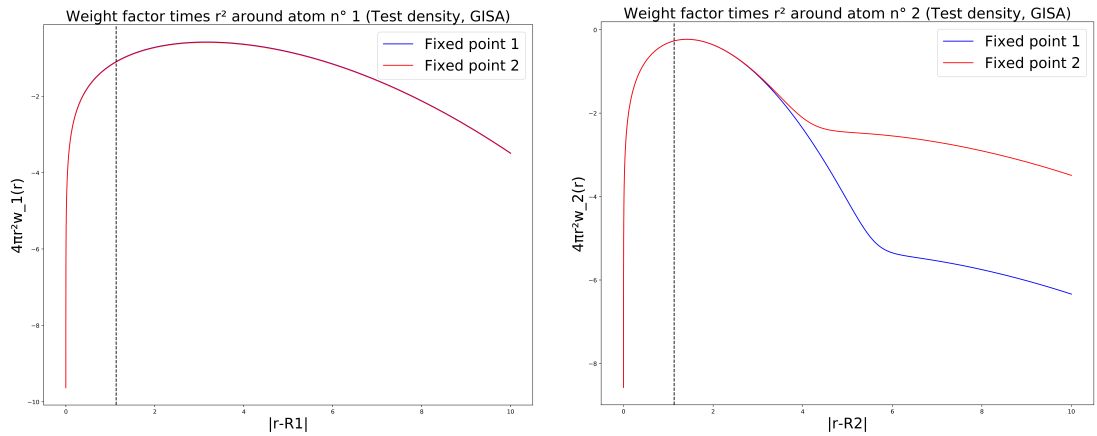

Figure 7: Pro-atomic density profiles $r \mapsto \log\left(4\pi r^2 \rho_a^{0,*}(r)\right)$ and $r \longmapsto \log\left(4\pi r^2 \rho_a^{0,**}(r)\right)$ associated to the two unveiled fixed points (the expected one, and another one) of the GISA algorithm.

Appendix B. Implicit function theorem for Lipschitz functions

Theorem 13. *[[63]] Let U and V open subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, $F \in C^{0,1}(U \times V; \mathbb{R}^n)$, and $(a,b) \in U \times V$ such that $F(a,b) = 0$. Assume that there exists $K > 0$ such that*

$$\forall (x,(y_1,y_2)) \in U \times (V \times V), \quad |F(x,y_1) - F(x,y_2)| \geq K|y_1 - y_2|.$$

Then, there exists an open neighborhood $\widetilde{U}$ of a in U and a Lipschitz function $\phi : \widetilde{U} \to V$ such that $\phi(a) = b$ and

$$\{(x,y) \in \widetilde{U} \times V \mid F(x,y) = 0\} = \{(x,\phi(x)),\, x \in \widetilde{U}\}.$$

In particular, $F(x,\phi(x)) = 0$ for all $x \in \widetilde{U}$.

References

[1] Carlo Adamo and Vincenzo Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, 1999.

[2] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. CVXOPT: A Python package for convex optimization.

[3] Paul W Ayers. Atoms in molecules, an axiomatic approach. i. maximum transferability. *The Journal of Chemical Physics*, 113(24):10886–10898, 2000.

[4] R. F. W. Bader and P. M. Beddall. Virial field relationship for molecular charge distributions and the spatial partitioning of molecular properties. *The Journal of Chemical Physics*, 56(7):3320–3329, 1972.

[5] Richard F. W. Bader. *Atoms in Molecules. A Quantum Theory.* Clarendon Press, 1994.

[6] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

[7] A. D. Becke. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of Chemical Physics*, 88(4):2547–2553, 1988.

[8] Robert Benda. *Modélisation multi-échelle de nano-capteurs à base de polymères conjugués pour la qualité de l'eau.* PhD thesis, Institut Polytechnique de Paris, 2021.

[9] Robert Benda, Eric Cancès, Virginie Ehrlacher, and Benjamin Stamm. Dma, https://github.com/rbenda/dma_multipoles.

[10] Patrick Bultinck, Paul W Ayers, Stijn Fias, Koen Tiels, and Christian Van Alsenoy. Uniqueness and basis set dependence of iterative hirshfeld charges. *Chemical physics letters*, 444(1-3):205–208, 2007.

[11] Patrick Bultinck, Christian Van Alsenoy, Paul W Ayers, and Ramon Carbó-Dorca. Critical analysis and extension of the hirshfeld atoms in molecules. *The Journal of chemical physics*, 126(14):144111, 2007.

[12] Christophe Chipot, Janos G Angyan, Gyorgy G Ferenczy, and Harold A Scheraga. Transferable net atomic charges from a distributed multipole analysis for the description of electrostatic properties: a case study of saturated hydrocarbons. *The Journal of Physical Chemistry*, 97(25):6628–6636, 1993.

[13] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[14] Martin S. Engler, Bertrand Caron, Lourens Veen, Daan P. Geerke, Alan E. Mark, and Gunnar W. Klau. Multiple-Choice Knapsack for Assigning Partial Atomic Charges in Drug-Like Molecules. In Laxmi Parida and Esko Ukkonen, editors, *18th International Workshop on Algorithms in Bioinformatics (WABI 2018)*, volume 113 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 16:1–16:13, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[15] György G Ferenczy. Charges derived from distributed multipole series. *Journal of computational chemistry*, 12(8):913–917, 1991.

[16] György G. Ferenczy, Peter J. Winn, and Christopher A. Reynolds. Toward improved force fields. 2. Effective distributed multipoles. *The Journal of Physical Chemistry A*, 101(30):5446–5455, 1997.

[17] Pierre Fortin. *Algorithmique hiérarchique parallèle haute performance pour les problèmes à N-corps*. PhD thesis, Université Sciences et Technologies-Bordeaux I, 2006.

[18] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T.ØSø rensen. The electron density is smooth away from the nuclei. *Communications in Mathematical Physics*, 228:401–415, 2002.

[19] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian 16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

[20] Donald Goldfarb and Ashok Idnani. Dual and primal-dual methods for solving strictly convex quadratic programs. In *Numerical analysis*, pages 226–239. Springer, 1982.

[21] Donald Goldfarb and Ashok Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming*, 27(1):1–33, 1983.

[22] Farnaz Heidar-Zadeh, Paul W Ayers, and Patrick Bultinck. Deriving the hirshfeld partitioning using distance metrics. *The Journal of chemical physics*, 141(9):094103, 2014.

[23] Farnaz Heidar-Zadeh, Paul W Ayers, Toon Verstraelen, Ivan Vinogradov, Esteban Vöhringer-Martinez, and Patrick Bultinck. Information-theoretic approaches to atoms-in-molecules: Hirshfeld family of partitioning schemes. *The Journal of Physical Chemistry A*, 122(17):4219–4245, 2017.

[24] F.L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica Chimica Acta*, 44(2):129–138, 1977.

[25] F.L. Hirshfeld. XVII. Spatial partitioning of charge density. *Israel Journal of Chemistry*, 16(2-3):198–201, 1977.

[26] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[27] Frank Jensen. *Introduction to computational chemistry*. John Wiley & Sons, 2017.

[28] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

[29] Vyacheslav Ivanovich Lebedev and DN Laikov. A quadrature formula for the sphere of the 131st algebraic order of accuracy. In *Doklady Mathematics*, volume 59, pages 477–481. Pleiades Publishing, Ltd., 1999.

[30] Timothy C Lillestolen and Richard J Wheatley. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chemical Communications*, 45:5909–5911, 2008.

[31] Timothy C. Lillestolen and Richard J. Wheatley. Atomic charge densities generated using an iterative stockholder procedure. *The Journal of Chemical Physics*, 131(14):144101, 2009.

[32] Chengwen Liu, Jean-Philip Piquemal, and Pengyu Ren. AMOEBA+ classical potential for modeling molecular interactions. *Journal of chemical theory and computation*, 15(7):4122–4139, 2019.

[33] Per-Olov Löwdin. Approximate formulas for many-center integrals in the theory of molecules and crystals. *The Journal of Chemical Physics*, 21(2):374–375, 1953.

[34] Turney Justin M, Simmonett Andrew C, Parrish Robert M, Hohenstein Edward G, Evangelista Francesco A, Fermann Justin T, Mintz Benjamin J, Burns Lori A, Wilke Jeremiah J, Abrams Micah L, et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(4):556–565, 2012.

[35] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[36] EN Maslen and MA Spackman. Atomic charges and electron density partitioning. *Australian journal of physics*, 38(3):273–288, 1985.

[37] Alston J. Misquitta, Anthony J. Stone, and Farhang Fazeli. Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. *Journal of Chemical Theory and Computation*, 10(12):5405–5418, 2014.

[38] R. S. Mulliken. Electronic population analysis on LCAO–MO molecular wave functions. i. *The Journal of Chemical Physics*, 23(10):1833–1840, 1955.

[39] Roman F Nalewajski and Robert G Parr. Information theory, atoms in molecules, and molecular similarity. *Proceedings of the National Academy of Sciences*, 97(16):8879–8882, 2000.

[40] Noel M. O'boyle, Adam L. Tenderholt, and Karol M. Langner. cclib: A library for package-independent computational chemistry algorithms. *Journal of Computational Chemistry*, 29(5):839–845, 2008.

[41] Robert G Parr, Paul W Ayers, and Roman F Nalewajski. What is an atom in a molecule? *The Journal of Physical Chemistry A*, 109(17):3957–3959, 2005.

[42] David A. Pearlman, David A. Case, James W. Caldwell, Wilson S. Ross, Thomas E. Cheatham, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis,

molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.

[43] Joshua A Rackers, Qiantao Wang, Chengwen Liu, Jean-Philip Piquemal, Pengyu Ren, and Jay W Ponder. An optimized charge penetration model for use with the AMOEBA force field. *Physical Chemistry Chemical Physics*, 19(1):276–291, 2017.

[44] Pengyu Ren and Jay W. Ponder. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry*, 23(16):1497–1506, 2002.

[45] Pengyu Ren and Jay W Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.

[46] Pengyu Ren, Chuanjie Wu, and Jay W. Ponder. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation*, 7(10):3143–3161, 2011.

[47] Christopher A. Reynolds, Jonathan W. Essex, and W. Graham Richards. Atomic charges for variable molecular conformations. *Journal of the American Chemical Society*, 114(23):9075–9079, 1992.

[48] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery. General atomic and molecular electronic structure system. *Journal of Computational Chemistry*, 14(11):1347–1363, 1993.

[49] Walter R. P. Scott, Philippe H. Hünenberger, Ilario G. Tironi, Alan E. Mark, Salomon R. Billeter, Jens Fennen, Andrew E. Torda, Thomas Huber, Peter Krüger, and Wilfred F. van Gunsteren. The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A*, 103(19):3596–3607, 1999.

[50] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren. Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of Chemical Theory and Computation*, 9(9):4046–4063, 2013.

[51] U. Chandra Singh and Peter A. Kollman. An approach to computing electrostatic charges for molecules. *Journal of Computational Chemistry*, 5(2):129–145, 1984.

[52] Anthony J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters*, 83(2):233–239, 1981.

[53] Anthony J Stone. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation*, 1(6):1128–1132, 2005.

[54] Anthony J. Stone and M. Alderton. Distributed multipole analysis. *Molecular Physics*, 56(5):1047–1064, 1985.

[55] Qiming Sun, Timothy C Berkelbach, Nick S Blunt, George H Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D McClain, Elvira R Sayfutyarova, Sandeep Sharma, et al. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018.

[56] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S Blunt, Nikolay A Bogdanov, George H Booth, Jia Chen, Zhi-Hao Cui, et al. Recent developments in the PySCF program package. *The Journal of chemical physics*, 153(2):024109, 2020.

[57] Toon Verstraelen, Paul W Ayers, Veronique Van Speybroeck, and Michel Waroquier. The conformational sensitivity of iterative stockholder partitioning schemes. *Chemical Physics Letters*, 545:138–143, 2012.

[58] Toon Verstraelen, Pawel Tecmer, Farnaz Heidar-Zadeh, Katharina Boguslawski, Matthew Chan, Yilin Zhao, Taewon D. Kim, Steven Vandenbrande, Derrick Yang, Cristina E. González-Espinoza, Stijn Fias, Peter A. Limacher, Diego Berrocal, Ali Malek, and Paul W. Ayers. HORTON 2.0.1, http://theochem.github.com/horton/, 2015.

[59] Toon Verstraelen, Steven Vandenbrande, Farnaz Heidar-Zadeh, Louis Vanduyfhuys, Veronique Van Speybroeck, Michel Waroquier, and Paul W Ayers. Minimal basis iterative stockholder: atoms in molecules for force-field development. *Journal of Chemical Theory and Computation*, 12(8):3894–3912, 2016.

[60] Fabienne Vigné-Maeder and Pierre Claverie. The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *The Journal of Chemical Physics*, 88(8):4934–4948, 1988.

[61] Peter J. Winn, György G. Ferenczy, and Christopher A. Reynolds. Toward improved force fields. 1. multipole-derived atomic charges. *The Journal of Physical Chemistry A*, 101(30):5437–5445, 1997.

[62] Johnny C. Wu, Gaurav Chattree, and Pengyu Ren. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theoretical Chemistry Accounts*, 131(3), 2012.

[63] Michael Wuertz. *The implicit function theorem for Lipschitz functions and applications.* PhD thesis, University of Missouri–Columbia, 2008.