Computational and Theoretical Chemistry 1074 (2015) 50–57

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

Direct minimization: Alternative to the traditional L_2 norm to derive partial atomic charges

Thomas Haschka [a,b], Eric Hénon [c,*], Christophe Jaillet [d], Laurent Martiny [a], Catherine Etchebest [b], Manuel Dauchez [a]

[a] MEDyC, SiRMa, CNRS UMR 7369, Université de Reims Champagne-Ardenne, BP 1039, 51687 Reims, France
[b] Dynamique des Systèmes et Interactions des Macromolécules Biologiques, INSERM UMR_S 1134, 6, rue Alexandre Cabanel, 75739 Cedex 15, Paris, France
[c] Institut de Chimie Moléculaire de Reims, CNRS UMR 7312, Université de Reims Champagne-Ardenne, BP 1039, 51687 Reims, France
[d] CReSTIC EA 3804, Université de Reims Champagne-Ardenne, BP 1039, 51687 Reims, France

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 8 October 2015
Accepted 8 October 2015
Available online 23 October 2015

Keywords:
Partial charges
L_1 metric
ESP fitting
Ab initio calculations
Minimization

ABSTRACT

Partial charges play an important role in simulating and understanding molecular properties. The derivation of an accurate charge model for monopolar atoms is a significant part of the parameterization of today's classical molecular mechanics force fields used in molecular dynamics simulations (MD). Hence, interest in the accurate prediction of partial charges from *ab initio* methods exists for a long time. Several methods have been developed, either based on population analysis that partitions the electron density into atomic populations, or on the assignment of partial atomic charges to reproduce a precalculated electrostatic potential (ESP method). In the latter approach, the charges are represented by parameters that are optimized by minimizing a loss function. ESP charge fitting, which is addressed in our work, is in most cases performed by minimization of a least squares or L_2 like loss functions. To our knowledge, no attempt was made to use different metrics such as least absolute deviations L_1 and to study their influence on the derived charges. The possibility of using different metrics to derive atomic charges is explored in this paper as a further extension of the ESP method. A direct iterative steepest descent minimization approach is employed in order to treat loss functions based on norms such as L_1. The implemented algorithm allows for dealing with chemical equivalency and total charge constraints while permitting using different loss functions. We compare the results from the L_1 norm to the values obtained from the standard L_2 norm and the L_4 norm for the 20 standard amino acids.

We suggest that close to the solution the L_1 norm expresses the impact of the electrostatic potential on the partial atomic charges to be obtained more accurately.

1. Introduction

Ab initio methods play an important role in defining parameters for mechanical force fields used to evaluate atomic interactions in molecular dynamics (MD) simulations of biomolecules. One of the crucial parameters derived from such methods is the atomic partial charge attributed to each atom, or cluster of objects, in coarse grained simulations, which is required for calculating force field coulombic part. Compared to the other force field terms, the coulomb term remains significant at long distance and is as such not limited to the immediate neighbors of the partial charge. Consequently, distance cutoffs, which are frequently used to speed up calculations, need to be larger for this term making it the most computational demanding term.

Among numerous methods developed during the last decades, three ways of charge attribution have become popular:

1. Population analysis, which puts an object's charge in direct relation to the probability of finding an electron in a certain region, orbital, next to the object.
   This method relies on integrating the electron density in different regions:
   - in real space,
   - or by basis set based partitioning according to the one particle density matrix.
     A well known real space partitioning methods is for instance Hirshfeld population analysis [1]. Popular basis set

---

* Corresponding author.

E-mail addresses: haschka@gmail.com (T. Haschka), eric.henon@univ-reims.fr (E. Hénon), christophe.jaillet@univ-reims.fr (C. Jaillet), laurent.martiny@univ-reims.fr (L. Martiny), catherine.etchebest@inserm.fr (C. Etchebest), manuel.dauchez@univ-reims.fr (M. Dauchez).

http://dx.doi.org/10.1016/j.comptc.2015.10.008
2210-271X/

partitioning methods are Mulliken analysis [2], "natural bond orbital" analysis (NBO) [3]. The further popular method "atom in molecules" (AIM) can be, [4,5] depending on its formulation, attributed to both groups. A complete overview of these methods is out of scope of this article, for divers effects on interesting chemical properties the reader shall be referred to Saha et al. [6].

2. Charge equilibration methods based on the QEq method originally developed by Rappé and Goddard [7]. This method works by equilibrating an electronegativity term χ_i for each atom i composing the molecule. These methods are based on several empirical values, but have the advantage, that once proper parameterization is achieved no further *ab initio* calculations are required. Hence, these charges are used where dynamic charge models are needed [8], taking the inherent inaccuracies into account.
3. Sampling the electrostatic potential derived from *ab initio* calculations and attributing charges in such a way that they reproduce the *ab initio* potential as good as possible (ESP method). [9–13].

Our purpose is not intended to discuss about the advantages and drawbacks of these methods. Rather, our work focuses on the widely used approach, i.e., the ESP method, a L_2 norm based approach. We have investigated alternative L_n norms that can be of interest. To set the context of our work, let's briefly sketch how this method is applied in general.

One essential ingredient is the "reference" electrostatic potential (ESP) $\Gamma(r)$ obtained at r by means of quantum mechanical calculations. The second is the predicted potential $\Phi(r)$ at point r which is obtained by assigning point charges Q_i to the atoms at positions x_i:

$$\Phi(r) = \sum_{i=1}^{\text{natoms}} \frac{Q_i(x_i)}{|r - x_i|}. \tag{1}$$

The quality of the potential generated by the charge model Φ in reference to Γ is in general expressed by a loss function F:

$$F = \sum_{i=1}^{\text{npoints}} D(\Gamma(p_i), \Phi(p_i)), \tag{2}$$

where D is a measure of the deviation at point p_i between the reference potential Γ and the potential Φ. In standard procedures that use efficient linear regression algorithms to look for the best possible solution, D is typically the squared difference between the two potential values at each sampling point p_i. A set of partial charges Q_i that minimizes the loss function F shall be our solution [11–14].

Beside least square deviations, we wondered whether alternative metrics could contribute to build a better set of partial atomic charges or not. This however, makes the commonly employed linear regression procedure used in L_2-ESP unfeasible. Thus we chose the gradient descent algorithm to minimize the loss function, as it is simple to implement and allows for a large number of metrics.

Modern computational methods to obtain atomic partial charges using the ESP approach account for additional factors such as chemical equivalency, rotational and conformational dependencies. To weaken these dependencies several point selection schemes have been proposed which are differing in the number of points distributed around the molecule and further in the way the points and their density are selected for the fit. Most methods implement either grid- or surface-based sampling point distributions [14]. The present study does not focus on these factors. However, in our program, we have paid special attention to account for these considerations (Section 2).

Two other points deserve to be highlighted in view of the results presented below. Firstly, weak coupling and thus collinearity frequently occur in the equations used in the standard ESP approach. The problem here is that the value of the electrostatic potential at point p_i in Eq. (2) might be only weakly influenced by the partial charge to be parameterized as the electrostatic potential diminishes distant from the molecule. A more general problem was pointed out by Francl et al. [15] The set of linear equations used to solve the ESP problem is, in the standard least squares fashion, already underdetermined for small molecules. This problem of collinearity further worsens as the molecules' size increases. Thus, in many cases, the least squares solution is not unique at a machines' precision and more than one set of partial charges satisfy the problem. In other words, a charge cannot be always meaningfully assigned to each atom, even for small molecules without buried atoms [15].

A second point concerns the shielding of buried charges. Charges assigned to atoms having almost no accessible surface area fall into this category (typically a carbon atom in a methyl group). Collinearity occurs in such cases. One can explain this by imagining a spherical homogeneous charge around a point charge. In such a case, infinite solutions can be found that recreate the potential surrounding such a system.

In the RESP approach [13] both problems are dealt with an iterative procedure arbitrarily confining buried charges around zero.

This paper is an attempt to use different metrics such as L_1 (least absolute deviations) and to study their influence on the derived charges. The adaptability of the metric D (Section 2) used for the loss function is the main feature of the approach presented herein. We show in Sections 3 and 4 that the L_1 metric, proven to be advantageous in various applications such as image recognition and machine learning, is at least as accurate as the L_2 based approach in deriving partial atomic charges and permanent dipole moments for a set of 20 amino acids. This can be explained, first, by improved sensitivity of the loss function to small changes in partial charges close to the optimal configuration and second, by enhanced statistics across the sample points p_i.

2. Methods

2.1. Implementation

2.1.1. Gradient descent

The approach presented herein relies on a given electrostatic reference potential evaluated on a point grid. As we selected a large number of sample points, we parallelized and implemented our algorithm in OpenCL.[1] A reference potential can be obtained by various *ab initio* software packages such as Gaussian [16] or NWChem [17]. Knowing the *ab initio* potential we can formulate a loss function (2) and minimize it by variation of the charges Q_i using a steepest descent algorithm.

A specific implementation of the traditional gradient descent algorithm is presented in Appendix A. The algorithm accounts for the standard chemical constraints, i.e. a total charge and chemical equivalency supplied by the user. Conceptual projection into a hyperplane in charge space is used to enforce these constraints. Apart from this specific feature, a conventional algorithm has been implemented.

2.1.2. Loss function

One of the key advantages of the gradient descent approach is that we can use almost any kind of metric D writing down our loss function (A.4). To go beyond the traditional square deviation, L_2

---

[1] The program is available at https://github.com/haschka/multicube.