1-1-2015

Electrostatic Point Charge Fitting as an Inverse Problem: Revealing the Underlying Ill-Conditioning

Maxim Vadimovich Ivanov
Marquette University

Marat R. Talipov
Marquette University, marat.talipov@marquette.edu

Qadir K. Timerghazin
Marquette University, qadir.timerghazin@marquette.edu

Electrostatic point charge fitting as an inverse problem: Revealing the underlying ill-conditioning

Maxim V. Ivanov, Marat R. Talipov, and Qadir K. Timerghazin[a)]
Department of Chemistry, Marquette University, Milwaukee, Wisconsin 53201-1881, USA

Atom-centered point charge (PC) model of the molecular electrostatics—a major workhorse of the atomistic biomolecular simulations—is usually parameterized by least-squares (LS) fitting of the point charge values to a reference electrostatic potential, a procedure that suffers from numerical instabilities due to the ill-conditioned nature of the LS problem. To reveal the origins of this ill-conditioning, we start with a general treatment of the point charge fitting problem as an inverse problem and construct an analytical model with the point charges spherically arranged according to Lebedev quadrature which is naturally suited for the inverse electrostatic problem. This analytical model is contrasted to the atom-centered point-charge model that can be viewed as an irregular quadrature poorly suited for the problem. This analysis shows that the numerical problems of the point charge fitting are due to the decay of the curvatures corresponding to the eigenvectors of LS sum Hessian matrix. In part, this ill-conditioning is intrinsic to the problem and is related to decreasing electrostatic contribution of the higher multipole moments, that are, in the case of Lebedev grid model, directly associated with the Hessian eigenvectors. For the atom-centered model, this association breaks down beyond the first few eigenvectors related to the high-curvature monopole and dipole terms; this leads to even wider spread-out of the Hessian curvature values. Using these insights, it is possible to alleviate the ill-conditioning of the LS point-charge fitting without introducing external restraints and/or constraints. Also, as the analytical Lebedev grid PC model proposed here can reproduce multipole moments up to a given rank, it may provide a promising alternative to including explicit multipole terms in a force field. © 2015 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4932105]

I. INTRODUCTION

The atom-centered point charge (PC) model of molecular electrostatics has been a mainstay of biomolecular simulations for decades.[1–14] While chemically intuitive and straightforward in technical implementation, this model does not provide a sufficiently detailed description of the anisotropic features of the molecular electrostatic potential (MEP), such as lone pairs, π-systems, and σ-holes, which are mostly governed by higher-order multipole terms.[15,16] These anisotropic effects, however, can be described within the PC approximation by moving beyond the atom-centered paradigm, i.e., by adding non-atom centered PCs/extended points.[17–20] Although increasing the number of PCs per atom improves the quality of the electrostatic model, it also can exacerbate well-known ill-conditioning and redundancy problems[21–23] of the PC fitting procedures, leading to numerically unstable solutions.[3,24,25]

These numerical instabilities are usually related to a large variation of the PC values for atoms in the interior of the molecule, so-called buried atom effect.[3,4,26,27] The buried atom (usually methyl and methylene carbons) charges can dramatically change due to trivial changes in the PC fitting problem (the probe grid sampling, spatial orientation of the molecule, etc.) and/or have inconsistent values across very similar molecules or even conformers of the same

molecule.[28,29] As the inclusion of non-atom centered PCs into the model produces even more buried centers, it should also increase the numerical instabilities of the PC fitting problem.

In fact, these numerical problems are rooted in the mathematical nature of the PC derivation—the least squares (LS) fitting to the reference MEP,[1,2]

$$\chi^2(\mathbf{q}) = |\mathbf{\Phi} - \mathbf{Aq}|^2 = |\mathbf{\Phi}|^2 + \mathbf{g}^\top \cdot \mathbf{q} + \mathbf{q}^\top \mathbf{Hq}, \quad (1)$$

$$\mathbf{g} = -2\mathbf{A}^\top\mathbf{\Phi}, \quad (2)$$

$$\mathbf{H} = \mathbf{A}^\top\mathbf{A}, \quad (3)$$

where the LS sum χ^2 is the subject of minimization and the solution satisfies normal equations,[30]

$$\mathbf{A}^\top\mathbf{Aq} = \mathbf{A}^\top\mathbf{\Phi}. \quad (4)$$

Here, the elements of the LS matrix $\mathbf{A}$ correspond to the inverse distance $1/r_{ij}$ between the PC i and the grid point j; $\mathbf{\Phi}$ is T-dimensional vector of the reference values of MEP; $\mathbf{q}$ is N-dimensional vector of the PC values; $\mathbf{g}$ is the gradient of the function χ^2 at the origin ($\mathbf{q} = 0$); $\mathbf{H}$ is the Hessian matrix of LS sum χ^2.

While the ill-conditioning is common to many LS fitting problems,[31–33] numerical difficulties associated with PC fitting are further compounded by commonly used total charge constraint using Lagrange multiplier.[21,22,34,35]

One of the most widely used techniques to alleviate the numerical instabilities of PC fitting is to add artificial

---

[a)]qadir.timerghazin@marquette.edu

restraints to the PC values of the buried atoms.[3,6,36,37] Although this method can be extended to models with off-center PCs/extended points, one may wonder if it would be possible to overcome these difficulties in a more elegant way, based on better physical understanding of the problem.

For instance, an important insight can be gleaned from the eigendecomposition of the LS sum Hessian matrix (Eq. (3)),

$$\mathbf{H}\mathbf{u}_i = \kappa_i \mathbf{u}_i. \tag{5}$$

Indeed, the ill-conditioned nature of the LS matrix $\mathbf{A}$ can be related to the significant differences in the eigenvalues κ_i, i.e., the LS sum curvatures along the directions defined by the eigenvectors $\mathbf{u}_i$.[38,39] Because of the 2–3 order of magnitude variation of the κ_i values, different sets of PCs can produce essentially the same MEP, as these solutions have the same positions along the high-curvature directions, although the positions along the low-curvature directions could be quite different.[38] Importantly, the eigenvectors with the largest curvatures usually correspond to the total charge and dipole moment components of the molecule, while the lower-curvature eigenvectors do not seem to be associated with particular multipole moments.[38,40]

However, the exact physical origin of the correspondence between the large curvature eigenvectors and the first terms of the multipole expansion is unclear, along with the nature of the low-curvature eigenvectors. Particularly, it is not clear if the presence of the low-curvature modes of the $\mathbf{H}$ matrix and thus the ill-conditioning of the LS problem is solely because of the nature of the PC fitting problem or due to some numerical factors, e.g., an incomplete sampling of the reference MEP grid.

To address these questions, we here revisit the PC fitting problem from the first principles. While the atom-centered PC model traces back to the intuitive chemical concept of the atomic charge, we consider a general PC model as a case of the inverse problem, where one seeks to recover the source charge distribution from its effect, i.e., electrostatic potential distribution. Based on the properties of the Coulomb law, we construct a best-case electrostatic model for which the inverse problem can be solved exactly, both in the continuous case and in the case of a discrete (non-atom centered) PC approximation.

Using this model, we investigate the nature of the eigenvectors $\mathbf{u}_i$ and their eigenvalues κ_i, dissect the factors responsible for the ill-conditioning of the LS fitting problem, and discuss how these insights can be used to improve and simplify the existing PC derivation procedures.

II. POINT CHARGE FITTING AS AN INVERSE PROBLEM

A problem where given an effect (in this case, the MEP Φ) defined in the region V_Φ, its cause (a charge distribution ρ) defined in the region V_ρ needs to be determined belongs to a general class of inverse problems and can be described by the Fredholm integral equation of the first kind,[41]

$$\int_{V_\rho} k(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')\, d\mathbf{r}' = \Phi(\mathbf{r}), \tag{6}$$

where kernel $k(\mathbf{r},\mathbf{r}')$ specifies the evolution of the cause $\rho(\mathbf{r}')$ into the effect $\Phi(\mathbf{r})$ that in this case corresponds to the Coulomb law,

$$k(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}. \tag{7}$$

The integral equation can also be represented as an operator equation,

$$K\rho = \Phi, \tag{8}$$

where $K : U \to V$ is a linear operator defined on space $U = \mathrm{range}(K^*) \in L^2$ of square integrable functions and takes values in space $V = \mathrm{range}(K) \in L^2$; $K^* : V \to U$ is adjoint of K. This equation can be solved exactly if and only if $\Phi \in V$. However, in general it is not the case, so a function ρ that minimizes the residual norm $|\Phi - K\rho|$ is considered as the LS solution and thus satisfies the normal equation,[41,42]

$$K^*K\rho = K^*\Phi. \tag{9}$$

This LS solution can be obtained as the linear combination of the basis vectors $u_i \in U$,[41]

$$\rho = K^\dagger\Phi = \sum_{i=1}^{\infty} \frac{\langle\Phi, v_i\rangle}{\mu_i} u_i, \tag{10}$$

where $K^\dagger$ is the Moore-Penrose inverse, μ_i is a singular value, v_i and u_i are left and right singular vectors, respectively, and the inner product $\langle\Phi, v_i\rangle$ is defined as

$$\langle\Phi, v_i\rangle = \int_{V_\Phi} \Phi(\mathbf{r})v_i(\mathbf{r})d\mathbf{r}. \tag{11}$$

The orthogonal bases $\{u_i\}_{i=1}^{\infty}$ and $\{v_i\}_{i=1}^{\infty}$ also form the eigenbases of K^*K and KK^* with eigenvalues μ_i^2,

$$K^*Ku_i = \mu_i^2 u_i, \tag{12}$$

$$KK^*v_i = \mu_i^2 v_i. \tag{13}$$

To obtain a numerical solution to the integral equation (Eq. (6)), the regions over which the MEP and charge distribution are defined are sampled using a numerical quadrature. Given N quadrature nodes over the charge distribution and T nodes over the MEP region, the integral equation is transformed into a system of T linear equations,

$$\mathbf{K}\mathbf{q} = \mathbf{\Phi}, \tag{14}$$

where the $T \times N$ matrix $\mathbf{K}$ is identical to the LS matrix $\mathbf{A}$ from Eq. (1) and contains the kernel elements k_{ij}, as this matrix originates from the kernel $k(\mathbf{r},\mathbf{r}')$ in the integral equation (Eq. (6)). It will be further referred to as $\mathbf{K}$ in order to highlight its mathematical origin.

Then, the PC value at the node i is the product of the charge density ρ_i and the quadrature weight w_i,

$$q_i = \rho_i w_i. \tag{15}$$

Since the number of the reference values T is usually larger than the number of the unknown PC values N, the system of linear equations is overdetermined. Then, a solution that minimizes the LS sum $\chi^2(\mathbf{q})$ (Eq. (1)) and satisfies normal equations (Eq. (4)) is considered as the numerical solution to the integral equation (Eq. (6)). This solution can be obtained using singular value decomposition (SVD) of matrix $\mathbf{K}$,[30,39,42]

$$\mathbf{q} = K^\dagger \Phi = \sum_{i=1}^{r} \frac{\Phi \cdot \mathbf{v}_i}{\mu_i} \mathbf{u}_i, \tag{16}$$

where $K^\dagger$ is the Moore-Penrose pseudoinverse; μ_i are singular values of matrix $\mathbf{K}$; vectors $\mathbf{v}_i$ and $\mathbf{u}_i$ are left and right singular vectors. If the rank r of matrix $\mathbf{K}$ is less than the dimension of $\mathbf{q}$ ($r < N$), then the matrix $\mathbf{K}$ is rank deficient.

Similarly to the continuous case (Eqs. (12) and (13)), the orthogonal bases $\{v_i\}_{i=0}^{r}$ and $\{u_i\}_{i=0}^{r}$ form eigenbases for $\mathbf{KK}^\mathsf{T}$ and $\mathbf{K}^\mathsf{T}\mathbf{K}$,

$$\mathbf{KK}^\mathsf{T}\mathbf{v}_i = \mu_i^2 \mathbf{v}_i, \tag{17}$$

$$\mathbf{K}^\mathsf{T}\mathbf{K}\mathbf{u}_i = \mu_i^2 \mathbf{u}_i, \tag{18}$$

where $\mathbf{K}^\mathsf{T}\mathbf{K}$ is also a Hessian matrix (Eq. (5)) and μ_i^2 is identical to its eigenvalue κ_i, which is the χ^2 curvature along the direction $\mathbf{u}_i$,[43]

$$\mu_i^2 = \kappa_i. \tag{19}$$

In many LS problems, PC fitting included, the singular values vary in a wide range, revealing the underlying ill-conditioning.[21,22,31,32] As a singular value μ_i is a denominator in the LS solution (Eq. (16)), the smaller the singular value, the larger the effect of the corresponding singular vector $\mathbf{u}_i$ on the LS solution. Thus, even small variations along $\mathbf{u}_i$ with small singular value lead to a significant variations of the LS solution, although these variations do not lead to significant change in the quality of the fit χ^2.[38] To understand the origins of the ill-conditioning in PC fitting, we next consider a system for which the inverse electrostatic problem can be analytically solved.

III. THE TWO-SPHERE MODEL

The Coulomb kernel (Eq. (7)) can be conveniently expanded in terms of spherical harmonics so the source $\mathbf{r}'$ and the observation $\mathbf{r}$ coordinates are separated but share the same origin,[44,45]

$$k(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} \frac{r_<^l}{r_>^{l+1}} Y_{lm}(\widehat{\mathbf{r}'})Y_{lm}(\widehat{\mathbf{r}}), \tag{20}$$

where $\widehat{\mathbf{r}} = \mathbf{r}/r$ denotes the unit vector defined by the polar φ and azimuthal θ angles; $r_<$ is the smaller and $r_>$ is the larger of r and r'; Y_{lm} are orthogonal real-value spherical harmonics,[46]

$$\int_S Y_{lm}(\widehat{\mathbf{r}})Y_{l'm'}(\widehat{\mathbf{r}})d\Omega = \delta_{ll'}\delta_{mm'}, \tag{21}$$

where $d\Omega$ is the differential of the solid angle.

Then, in the region beyond the divergence sphere where the charge density vanishes, the MEP can be expanded in a multipole series,[44,45]

$$\Phi(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sqrt{\frac{4\pi}{2l+1}} r^{-l-1} Q_{lm}^{mol} Y_{lm}(\widehat{\mathbf{r}}), \tag{22}$$

where a molecular multipole moment Q_{lm}^{mol} is given by

$$Q_{lm}^{mol} = \sqrt{\frac{4\pi}{2l+1}} \int r^l \rho(\mathbf{r})Y_{lm}(\widehat{\mathbf{r}})d^3r. \tag{23}$$

The form of the kernel expansion (Eq. (20)) suggests that if the radii $r = R$ and $r' = a$ are fixed, the kernel $k(\mathbf{R},\mathbf{a})$ can uniquely map a charge density over a spherical surface S_a to the corresponding potential $\Phi(\mathbf{R})$ on a sphere S_R and vice versa. Thus, for a probe sphere S_R with the radius R greater than the radius of divergence sphere, the MEP can be reproduced exactly by a sphere S_a with surface charge density $\sigma(\mathbf{a})$ such that the multipole moments of the sphere $Q_{lm}^{S_a}$ are equivalent to the multipole moments of the molecule Q_{lm}^{mol},

$$Q_{lm}^{S_a} \equiv Q_{lm}^{mol}, \tag{24}$$

where multipole moments of the sphere are

$$Q_{lm}^{S_a} = \sqrt{\frac{4\pi}{2l+1}} a^l \int_{S_a} \sigma(\mathbf{a})Y_{lm}(\widehat{\mathbf{a}})d\Omega. \tag{25}$$

In this case, original integral Eq. (6) is transformed into a surface integral equation,

$$\int_{S_a} k(\mathbf{R},\mathbf{a})\sigma(\mathbf{a})d\Omega = \Phi(\mathbf{R}), \tag{26}$$

or, equivalently, in an operator form,

$$K\sigma = \Phi, \tag{27}$$

where $K : L^2(S_a) \to L^2(S_R)$ is a compact infinite-rank operator (Fig. 1(a)),

$$K\sigma = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \mu_l \langle \sigma, Y_{lm}^{S_a} \rangle Y_{lm}^{S_R}, \tag{28}$$

where subscripts S_a and S_R denote the spheres, on which the corresponding spherical harmonics are defined; the projection $\langle \sigma, Y_{lm}^{S_a} \rangle$ is the inner product on the $L^2(S_a)$ space,

$$\langle \sigma, Y_{lm}^{S_a} \rangle = \int_{S_a} \sigma(\mathbf{a})Y_{lm}^{S_a}(\widehat{\mathbf{a}})d\Omega \tag{29}$$

and for each degree l, there is a singular value μ_l in the form of the distance-dependent factor from the MEP expansion (Eq. (20)),

$$\mu_l = \frac{4\pi}{2l+1} \frac{a^l}{R^{l+1}}. \tag{30}$$

Accordingly, the spherical harmonics $Y_{lm}^{S_R}$ and $Y_{lm}^{S_a}$ are left and right singular vectors and thus the eigenfunctions of the operators K^*K and KK^*, while the squares of the singular values μ_l are their eigenvalues (Eqs. (12) and (13)). Since the
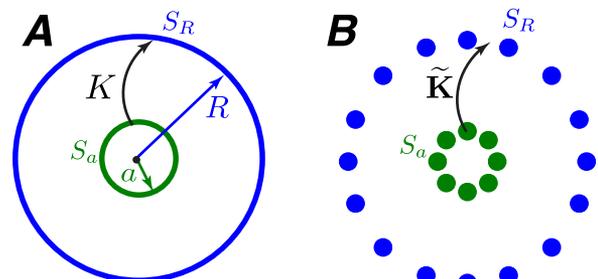

FIG. 1. Schematic representations of the probe S_R and charged S_a spheres in the continuous (a) and discrete (b) forms. Operators K (Eq. (27)) and $\widetilde{\mathbf{K}}$ (Eq. (39)) are represented schematically.

singular values μ_l and spherical harmonics $Y_{lm}^{S_a}$ and $Y_{lm}^{S_R}$ form a singular system of the operator K, the solution to integral equation (Eq. (26)) can be expressed as

$$\sigma = K^\dagger \Phi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{\langle \Phi, Y_{lm}^{S_R} \rangle}{\mu_l} Y_{lm}^{S_a}. \quad (31)$$

According to the multipole expansion (Eq. (22)), the inner product $\langle \Phi, Y_{lm}^{S_R} \rangle$ depends on the radius R of the probe sphere and the multipole moments of the molecule,

$$\langle \Phi, Y_{lm}^{S_R} \rangle = \int_{S_R} \Phi(\mathbf{R}) Y_{lm}^{S_R}(\hat{\mathbf{R}}) d\Omega = \sqrt{\frac{4\pi}{2l+1}} \frac{1}{R^{l+1}} Q_{lm}^{mol}. \quad (32)$$

The dependence on the radius R cancels out, so the charge density depends only on the radius a of the sphere S_a and the molecular multipole moments,

$$\sigma(\mathbf{a}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} a^{-l} Y_{lm}^{S_a}(\hat{\mathbf{a}}) Q_{lm}^{mol}, \quad (33)$$

and the charged sphere S_a exactly reproduces the MEP $\Phi(\mathbf{R})$.

IV. ANALYTICAL LEBEDEV GRID POINT CHARGE MODEL

We can construct an approximate discrete analog of the two-sphere model (Eqs. (26)-(33), Fig. 1) using a quadrature that exactly integrates spherical harmonics Y_{lm} over a sphere up to a given l (Eqs. (29) and (32)), e.g., the widely used[47–49] Lebedev quadrature,[50] that defines N quadrature nodes (Table 1 in the supplementary material[51]) with predetermined angular coordinates θ_i, φ_i, and integration weights w_i,

$$\int_S Y_{lm}(\theta, \varphi) d\Omega = \sum_{i}^{N} Y_{lm}(\theta_i, \varphi_i) w_i. \quad (34)$$

Then, given the surface charge density σ_i, the corresponding point charge is

$$q_i = \sigma_i w_i. \quad (35)$$

Due to the orthogonality of the spherical harmonics Y_{lm} (Eq. (21)), the N-node Lebedev quadrature that exactly integrates spherical harmonics over the sphere S_a up to $l = 2n$,

$$\sum_{i}^{N} Y_{lm}^{S_a}(\theta_i, \varphi_i) Y_{l'm'}^{S_a}(\theta_i, \varphi_i) w_i^{S_a} = \widetilde{\mathbf{Y}}_{lm}^{S_a} \cdot \widetilde{\mathbf{Y}}_{l'm'}^{S_a} = \delta_{ll'} \delta_{mm'}, \quad (36)$$

defines an orthonormal basis of dimension $d_n = (n+1)^2$,

$$\widetilde{\mathbf{Y}}_{S_a} = \{ \widetilde{\mathbf{Y}}_{lm}^{S_a}, \ -l \le m \le l \}_{l=0}^{n}, \quad (37)$$

where the $\widetilde{\mathbf{Y}}_{lm}^{S_a}$ vectors have N elements and are defined as

$$\widetilde{Y}_{lmi}^{S_a} = Y_{lm}(\theta_i, \varphi_i) \sqrt{w_i^{S_a}}. \quad (38)$$

Similarly, the probe sphere S_R can be represented by a T-node Lebedev grid that integrates spherical harmonics up to $l = 2t$ and defines an orthogonal basis $\widetilde{\mathbf{Y}}_{S_R}$ of dimension $d_t = (t+1)^2$.

In this discrete representation, the operator K (Eq. (28)) then becomes a $T \times N$ matrix $\widetilde{\mathbf{K}}$,[52]

$$\widetilde{\mathbf{K}} \widetilde{\sigma} = \widetilde{\mathbf{\Phi}}, \quad (39)$$

where the elements of $\widetilde{\mathbf{K}}$, $\widetilde{\sigma}$, and $\widetilde{\mathbf{\Phi}}$ are

$$\widetilde{K}_{ij} = \sqrt{w_i^{S_a} w_j^{S_R}} / r_{ij}, \quad (40)$$

$$\widetilde{\sigma}_i = \sigma_i \sqrt{w_i^{S_a}}, \quad \widetilde{\Phi}_j = \Phi_j \sqrt{w_j^{S_R}}. \quad (41)$$

Since usually the probe grid has more points than the source grid, i.e., $T > N$, the matrix equation (Eq. (39)) is a LS problem (Eq. (1)) that can be solved using SVD of the matrix $\widetilde{\mathbf{K}}$ (Eq. (16)), giving a discrete analog of Eq. (31),

$$\widetilde{\sigma} = \sum_{l=0}^{n} \sum_{m=-l}^{l} \frac{\widetilde{\mathbf{\Phi}} \cdot \widetilde{\mathbf{Y}}_{lm}^{S_R}}{\mu_l} \widetilde{\mathbf{Y}}_{lm}^{S_a}, \quad (42)$$

where $\widetilde{\mathbf{Y}}_{lm}^{S_R}$ and $\widetilde{\mathbf{Y}}_{lm}^{S_a}$ are left and right singular vectors, and the corresponding singular values μ_l are the same as in the continuous case (Eq. (30)).

Since we use the Lebedev quadrature, the dot product $\widetilde{\mathbf{\Phi}} \cdot \widetilde{\mathbf{Y}}_{lm}^{S_R}$ corresponds to exact numerical integration and gives a result identical with the continuous case (Eq. (32)),

$$\widetilde{\mathbf{\Phi}} \cdot \widetilde{\mathbf{Y}}_{lm}^{S_R} = \sum_{j=0}^{T} \Phi_j Y_{lmj}^{S_R} w_j = \sqrt{\frac{4\pi}{2l+1}} \frac{1}{R^{l+1}} Q_{lm}^{mol}, \quad (43)$$

so the solution to Eq. (39) depends only on the radius a and the multipole moments Q_{lm}^{mol},

$$\widetilde{\sigma} = \sum_{l=0}^{n} \sum_{m=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} a^{-l} Q_{lm}^{mol} \widetilde{\mathbf{Y}}_{lm}^{S_a}. \quad (44)$$

The corresponding PC values q_i can be obtained using the quadrature weights $w_i^{S_a}$,

$$q_i = \sigma_i w_i^{S_a} = \widetilde{\sigma}_i \sqrt{w_i^{S_a}}, \quad (45)$$

or, in a vector form,

$$\mathbf{q} = \sum_{l=0}^{n} \sum_{m=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} a^{-l} Q_{lm}^{mol} \mathbf{Y}_{lm}^{S_a} \odot \mathbf{w}^{S_a}, \quad (46)$$

where $\mathbf{w}^{S_a}$ is the vector of the quadrature weights for the sphere S_a. Therefore, we can use Lebedev grid that shares the origin with a molecule to construct an analytical PC model that exactly reproduces molecular multipole values up to the degree n.

From this model, we can see that the ill-conditioning of the PC fitting due to the decay of the singular values is intrinsic to the inverse electrostatic problem, as the singular values μ_l decrease with increasing l (Eq. (30)). Indeed, the higher the multipole moment, the smaller its contribution to the overall electrostatic potential. Also, this contribution gets smaller as we move the probe further away from the source, and the singular values get smaller with the increasing radius of the probe sphere R or decreasing radius of the source sphere a.

The ill-conditioning problems become even more severe as we switch from modeling the MEP using the Lebedev quadrature, which is best suited to reproduce the molecular
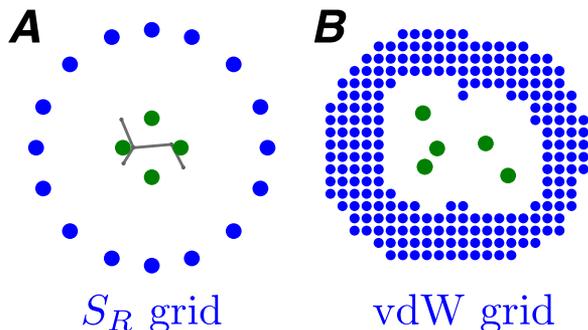
FIG. 2. Cross-sectional representations of the quadratures used for two-sphere model (a) ($n = 1$, $N = 6$ and $t = 11$, $T = 194$ for spheres S_a and S_R, respectively) as compared with the traditional atom-centered model (b). Green circles correspond to the point charges; blue circles correspond to the reference grid points.

multipoles, to an irregular atom-centered quadrature, as shown on a numerical example below.

V. LEBEDEV GRID VS. ATOM-CENTERED MODEL: A NUMERICAL EXAMPLE

First, we consider an electrostatic PC model of a methanol molecule with PCs placed at the nodes of the Lebedev quadrature over the sphere S_a ($a = 2$ a.u.) (Fig. 2(a)). In this case, the PC values can be obtained analytically from the reference multipole moments (Eq. (46)) or by numerical fitting to the reference MEP over the probe sphere S_R ($R = 8$ a.u., $t = 11$, $T = 194$),

$$\widetilde{\sigma} = \sum_{i=1}^{r} \frac{\widetilde{\mathbf{\Phi}} \cdot \widetilde{\mathbf{v}}_i}{\mu_i} \widetilde{\mathbf{u}}_i, \qquad (47)$$

where the PC value can be found as $q_j = \widetilde{\sigma}_j \sqrt{w_j^{S_a}}$ and the maximum rank r is the number N of quadrature nodes/PCs over the sphere S_a. The quality of the fit is measured using the root mean square deviation (RMSD) calculated over the T

nodes of the probe grid,

$$\text{RMSD} = \sqrt{\frac{\chi^2}{T}}. \qquad (48)$$

Naturally, the analytical PC values from Eq. (46) exactly reproduce the molecular multipole moments up to the degree n defined by the quadrature (Table I). For each degree l, there are $2l + 1$ values of order m, so overall $(n + 1)^2$ multipole moments are reproduced, which matches the dimension d_n of the corresponding basis $\widetilde{\mathbf{Y}}_{S_a}$ (Eq. (37)). As the dimension d_n increases, more multipole moments are reproduced and the RMSD rapidly approaches zero (Fig. 4 in the supplementary material[51]).

Since the dimension d_n does not match the number of quadrature nodes N (Table 1 in the supplementary material[51]),[53,54] we can obtain numerical solutions with Eq. (47) that are equivalent to the analytical results (Eq. (46)) by setting the rank r to the dimension of the grid, $d_n = (n + 1)^2$ (Table I).[55]

As the first d_n multipole moments Q_{lm}^{mol} are reproduced by the PC model, the first d_n numerical singular values μ_i exactly match the radius-dependent part (Eq. (30)) from the inverse distance expansion (Fig. 3), and the corresponding right singular vectors $\widetilde{\mathbf{u}}_i$ match the basis $\widetilde{\mathbf{Y}}_{S_a}$ (Fig. 4),

$$\{\widetilde{\mathbf{u}}_i\}_{i=1}^{d_n} = \{\widetilde{\mathbf{Y}}_{lm}^{S_a}, \ -l \le m \le l\}_{l=0}^{n}. \qquad (49)$$

If we do not restrict the rank r to the dimension of the grid d_n, numerical SVD of the LS matrix $\widetilde{\mathbf{K}}$ (Eq. (47)) produces N singular vectors/values. While this slightly improves the RMSD (Table I), the additional $N - d_n$ singular vectors cannot be described analytically (Fig. 4) as they go beyond the dimension d_n of the corresponding basis $\widetilde{\mathbf{Y}}_{S_a}$. However, in the fortuitous case of the quadrature with $n = 1$ and $N = 6$, the remaining $6 - 4 = 2$ vectors resemble the basis vectors $\widetilde{\mathbf{Y}}_{2-2}$ and $\widetilde{\mathbf{Y}}_{2-1}$, so the corresponding quadrupole moments Q_{2-2} and Q_{2-1} are accurately reproduced, although the exact numerical integration of the spherical harmonics Y_{2-2} and Y_{2-1} is not provided by the 6-node Lebedev grid.

TABLE I. Effect of the rank r (Eq. (47)), degree n (Eq. (46)), and type of the charge constraint on the methanol multipole moments and the RMSD (kcal/mol) within the Lebedev grid PC model ($a = 2$ a.u., $n = 1, 2$, and $N = 6, 14$) with probe sphere S_R ($R = 8$ a.u., $T = 194$) and atom-centered PC model with vdW-type grid.

| PC model, probe grid	Details	Q_0	Q_{10}	Q_{11}	Q_{1-1}	Q_{20}	Q_{21}	Q_{2-1}	Q_{22}	Q_{2-2}	RMSD
	Eq. (46), $n = 1$	0.000	0.000	−0.325	−0.565	0.000	0.000	0.000	0.000	0.000	1.762
$S_a(N = 6)$, $S_R(T = 194)$	SVD, $r = 4$	0.000	0.000	−0.323	−0.562	0.000	0.000	0.000	0.000	0.000	1.762
	SVD, $r = 6$	0.000	0.000	−0.323	−0.562	−0.881	0.000	0.000	−0.497	0.000	1.668
	Eq. (46), $n = 2$	0.000	0.000	−0.325	−0.565	−0.887	0.000	0.000	−0.503	2.882	0.559
$S_a(N = 14)$, $S_R(T = 194)$	SVD, $r = 9$	0.000	0.000	−0.325	−0.566	−0.881	0.000	0.000	−0.497	2.865	0.559
	SVD, $r = 14$	0.000	0.000	−0.325	−0.566	−0.881	0.000	0.000	−0.497	2.865	0.425
	SVD, $r = 6$	0.007	0.000	−0.335	−0.546	−1.020	0.001	0.001	0.110	2.761	2.457
	tSVD, $r = 5$	0.009	0.000	−0.338	−0.543	−1.060	−0.002	0.000	0.219	2.688	2.625
	tSVD, $r = 4$	0.003	0.001	−0.270	−0.377	0.574	−0.003	−0.003	0.225	−1.317	8.729
Atom-centered, vdW	Lagrange, $Q_0 = 0$	0.000	0.000	−0.332	−0.543	−0.989	0.001	0.001	0.073	2.739	2.587
	Elimination, $Q_0 = 0$	0.000	0.000	−0.332	−0.543	−0.989	0.001	0.001	0.073	2.739	2.587
	SVD, $Q_0 = 0$	0.000	0.000	−0.331	−0.544	−1.010	0.001	0.001	0.097	2.756	2.597
	Trivial, $Q_0 = 0$	0.000	0.000	−0.331	−0.544	−1.010	0.001	0.001	0.097	2.755	2.597
Reference		0.000	0.000	−0.325	−0.565	−0.887	0.000	0.000	−0.503	2.882	...
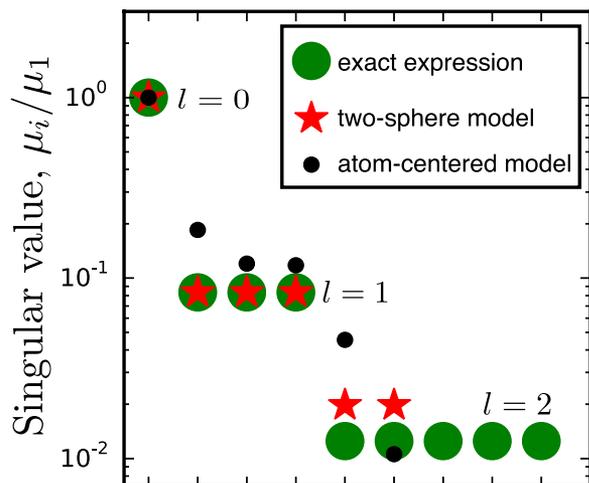
FIG. 3. Normalized singular values μ_i/μ_1 obtained using exact analytical expression Eq. (30) (green circles) as compared with the numerical values obtained from SVD of the LS matrix for the two-sphere model (red stars) and for atom-centered model (black circles). Lebedev quadratures with $n = 1$, $N = 6$ and $t = 11$, $T = 194$ were used for the charged S_a ($a = 2$ a.u.) and probe S_R ($R = 8$ a.u.) spheres, respectively.

Now, we can use the insights from the best-case scenario spherical PC model based on the Lebedev quadrature (Fig. 2(a)) to understand the traditional PC fitting problem with atom-centered charges and the probe grid that follows the solvent-accessible surface (van der Waals (vdW) grid, Fig. 2(b)). From the point of view of the inverse electrostatic model, the atom-centered PC fitting corresponds to a numerical solution using an irregular and suboptimal integration grid to represent the source charge distribution. This problem can be treated by SVD of the LS matrix $\mathbf{K}$,

$$\mathbf{q} = \sum_{i=1}^{r} \frac{\mathbf{\Phi} \cdot \mathbf{v}_i}{\mu_i} \mathbf{u}_i, \tag{50}$$

where the maximum value of rank r is the number of atoms in the molecule, i.e., $r = 6$ in the case of methanol.

We can see that even in this case, the singular vector $\mathbf{u}_1$ with the largest singular value μ_1 corresponds the total charge (Fig. 4), which is reproduced with only a slight numerical deviation (<0.01), a consequence of the molecular charge density spillover beyond the solvent-accessible surface defining the vdW grid.[38,40]

Although the other singular vectors do not exactly match the corresponding spherical harmonics, the $\mathbf{u}_2$–$\mathbf{u}_4$ vectors can

TABLE II. Effect of the numerical rank r (SVD in Eq. (50)) and the total-charge constraint on the values of atom-centered PCs of methanol and the RMSD (kcal/mol).

	q_C	q_{H_g}	q_{H_t}	q_O	q_H	RMSD
SVD, $r = 6$	0.215	−0.018	0.048	−0.592	0.371	2.457
tSVD, $r = 5$	−0.058	0.056	0.118	−0.532	0.370	2.625
tSVD, $r = 4$	0.007	0.089	−0.101	−0.070	−0.010	8.729
Lagrange, $Q_0 = 0$	0.276	−0.035	0.030	−0.603	0.367	2.587
Elimination, $Q_0 = 0$	0.276	−0.035	0.030	−0.603	0.367	2.587
SVD, $Q_0 = 0$	0.214	−0.019	0.047	−0.593	0.370	2.597
Trivial, $Q_0 = 0$	0.214	−0.019	0.047	−0.593	0.370	2.597

be roughly related to the three components of the dipole moment (Fig. 4), and the corresponding singular values are commensurate with the singular value μ_l ($l = 1$) obtained for the Lebedev grid model (Fig. 3). The remaining singular values μ_5 and μ_6 are significantly distorted from the singular value μ_l ($l = 2$), so the components of the quadrupole moment are not reproduced as precisely as the dipole moment components (Table I).

Among all singular vectors $\{\mathbf{u}_i\}_{i=1}^{6}$, the singular vector $\mathbf{u}_6$ with the lowest singular value μ_6, which is 100 times smaller than μ_1, is dominated by the contribution from the methyl carbon atom (Fig. 4). Since such small singular values cause numerical instabilities of the LS solution, one can use a regularization technique such as truncated SVD (tSVD) that reduces the rank r by removing the lowest-μ_i vector(s) from the SVD expansion.[39] Removal of $\mathbf{u}_6$ that decreases the rank to $r = 5$ leads to dramatic change in the methyl group charges— the carbon atom charge in particular, which drops from 0.22 to −0.06. Yet, these changes lead only to marginal changes in the multipole moment and RMSD values, a typical example of the buried atom effect (Tables I and II). This suggests a natural way to impose a restraint on the buried atom charges without introducing a restraining function into the LS sum χ^2, an addition that can negatively affect the electrostatic properties of the PC model.[22,56]

Further removal of the singular vectors $\mathbf{u}_5$ and $\mathbf{u}_6$ (i.e., $r = 4$) leads to severe deterioration of the LS solution, as the corresponding multipole moment strongly deviates from the reference values and the RMSD significantly increases (Tables I and II). Thus, it appears that the tSVD approach should be applied only to the singular vectors that strongly depend on the buried atoms, an important point that will be discussed in detail elsewhere.
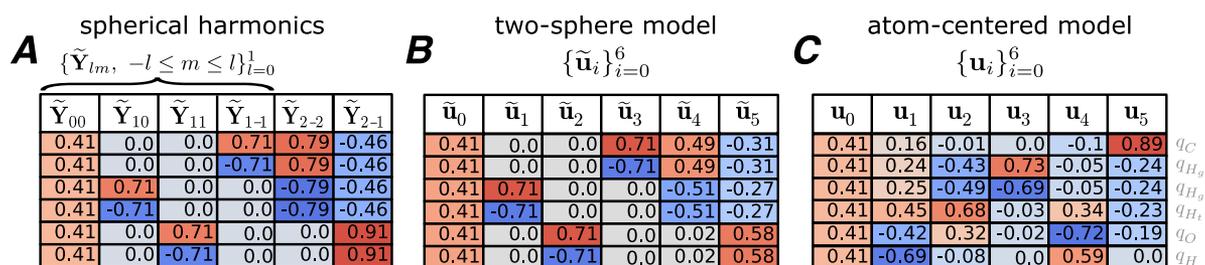

FIG. 4. The orthonormal bases of the right singular vectors: basis of spherical harmonics $\widetilde{\mathbf{Y}}_{S_a}$ (a), basis from the numerical SVD of the LS matrix in two-sphere PC model (b), and atom-centered model (c).

VI. THE TOTAL-CHARGE CONSTRAINT REVISITED

Commonly used PC fitting approaches also modify the LS sum (Eq. (1)) by adding a Lagrange multiplier λ in order to constrain the total charge to the correct value,[21,34,35]

$$\chi^2(\mathbf{q}) = |\mathbf{\Phi} - \mathbf{Kq}|^2 + \lambda(\mathbf{1}^\top \cdot \mathbf{q} - Q_0^{mol}), \quad (51)$$

which increases the dimension of the Hessian matrix $\mathbf{H} = \mathbf{K}^\top\mathbf{K}$ in the normal equation (Eq. (4)),

$$\begin{bmatrix} \mathbf{H} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{K}^\top\mathbf{\Phi} \\ Q_0^{mol} \end{bmatrix}, \quad (52)$$

where $\mathbf{1}$ is an all-ones column-vector.

However, as we have seen, both in the case of the Lebedev grid and the less-than-ideal atom-centered PC models, the Hessian eigenvector with the largest curvature corresponds to the total charge (Fig. 4 and also Ref. 38). Thus, in the case of the two-sphere PC fitting, the total charge is reproduced exactly ($Q_0 < 10^{-5}$), while in the atom-centered PC model, the total charge only slightly deviates from the exact value due the close proximity of the vdW grid and slight distortion of the total-charge vector $\mathbf{u}_1$ from its analytical analog $\widetilde{\mathbf{Y}}_0$ ($Q_0 = 0.003$ for methanol, Table I).

Addition of the Lagrange multiplier leads to an extra eigenvector $\mathbf{u}_7$ that appears in the eigenbasis of the Hessian matrix (Table VIII in the supplementary material[51]). The curvature along this vector is the smallest in the magnitude ($\kappa_7 = -0.009$) and the vector itself primarily depends on the Lagrange multiplier λ, with only marginal contribution from the PC values. At the same time, remaining eigenvectors $\{\mathbf{u}\}_{i=1}^6$ preserve the structure of the original eigenbasis, with negligible contribution from the Lagrange multiplier λ (Table VIII in the supplementary material[51]). Thus, application of the total charge constraint in addition to already strong restraint (imposed by the eigenvector $\mathbf{u}_1$) appears to be redundant. Moreover, addition of the Lagrange multiplier aggravates the rank deficiency of already ill-conditioned LS problem.[21,22]

Alternatively, the total charge can be constrained by incorporating condition on the proper total charge directly into the LS sum,[1,4,23] by eliminating one of the charges and setting it to

$$q_n = Q_0^{mol} - \sum_{i}^{N-1} q_i, \quad (53)$$

where n is the index of the eliminated charge. This reduces the dimension of the LS problem by one

$$\chi^2(\mathbf{q}) = \sum_{j}^{T} \left[\Phi_j - \frac{Q_0^{mol}}{r_{nj}} - \sum_{i}^{N-1} \left(\frac{1}{r_{ij}} - \frac{1}{r_{nj}} \right) q_i \right]^2 \quad (54)$$

and modifies the elements of the Hessian matrix,

$$H_{km} = \sum_{j}^{T} \left(\frac{1}{r_{kj}} - \frac{1}{r_{nj}} \right) \left(\frac{1}{r_{mj}} - \frac{1}{r_{nj}} \right). \quad (55)$$

Although the solution obtained with this approach is numerically equivalent to the solution with Lagrange multiplier, regardless which atom has been eliminated (elimination, $Q_0 = 0$ in Tables I and II), the structure of the right singular

vectors becomes disrupted (Fig. 8 in the supplementary material[51]), which prevents the application of the truncated SVD to improve the numerical stability of the solution.

Given that even for the atom-centered PC/vdW probe model the total charge value deviates only very slightly from the reference value, it should be possible to correct for this deviation without exacerbating the numerical instabilities of the LS problem, e.g., using the total charge vector $\mathbf{u}_1$. To do that, we convert the SVD solution (Eq. (50)) to a system of linear equations,

$$\mathbf{q} = \sum_{i=0}^{r} \underbrace{\frac{\mathbf{\Phi} \cdot \mathbf{v}_i}{\mu_i}}_{c_i} \mathbf{u}_i = \mathbf{Uc}, \quad (56)$$

$$\mathbf{U}^\top\mathbf{q} = \mathbf{c}. \quad (57)$$

Then, we replace $\mathbf{u}_1$ in $\mathbf{U}^\top$ by an all-ones vector $\mathbf{1}$ and set the corresponding coefficient c_1 in $\mathbf{c}$ to the exact value of the molecular total charge Q_0^{mol},

$$\mathbf{U}_{Q_0}^\top\mathbf{q} = \mathbf{c}_{Q_0}, \quad (58)$$

where

$$\mathbf{U}_{Q_0}^\top = \begin{bmatrix} \mathbf{1} & \mathbf{u}_2 & \cdots & \mathbf{u}_N \end{bmatrix}^\top, \quad (59)$$

$$\mathbf{c}_{Q_0} = \begin{bmatrix} Q_0^{mol} & c_2 & \cdots & c_N \end{bmatrix}^\top. \quad (60)$$

This approach does not introduce any redundant constraints, preserves the electrostatic properties of the unconstrained solution, and results only into minor changes in the PC values (SVD, $Q_0 = 0$ in Tables I and II) and is compatible with truncated SVD. Also, the error in the total charge value is small enough and can be corrected by simply distributing the Q_0 error correction across the atomic charges; this trivial total charge correction gives result nearly identical to Eq. (58) (trivial, $Q_0 = 0$ in Tables I and II).

VII. CONCLUSIONS

To understand the origins of the ill-conditioning of the LS PC fitting problem, we revisited the PC representation of the MEP from the first principles, as an example of the inverse problem.

Based on the properties of the Coulomb potential that can be expanded in terms of spherical harmonics, we introduce a model where the MEP of a molecule is exactly reproduced by a charged sphere that has the same multipole moments Q_{lm} as the molecule. Using Lebedev quadrature, this continuous model is converted into a discrete PC model, where the PC values are evaluated analytically from the multipole moments Q_{lm} up to the maximum value determined by the quadrature.

In this context, the traditional atom-centered PC model can be viewed as an irregular numerical quadrature, poorly suited to reproduce the multipolar expansion of the MEP. As such, this quadrature only allows integration of the monopole and, approximately, dipole terms. The corresponding large-curvature—or "stiff"[32,33]—Hessian eigenvectors $\mathbf{u}_i$ can still be related to the corresponding multipoles Q_{lm}. This explains previously observed correspondence between the highest-curvature Hessian eigenvectors and the total charge and the

dipole moment components;[38,40] this correspondence quickly breaks down for the higher multipole moments.

This consideration then reveals the origins of the ill-conditioning of the PC fitting due to the presence of low-curvature—or "sloppy"[32,33]—vectors $\mathbf{u}_i$. The intrinsic ill-conditioning arises even in the case of the ideal spherical model: since the higher-rank multipole moments Q_{lm} have smaller contribution to the MEP, the singular values μ_l decay as l increases. The ill-conditioning is further exacerbated in the numerical treatment of the Lebedev grid model because the number of PCs does not match the dimension of the basis formed by Lebedev quadrature. The remaining singular values/curvatures are even lower in magnitude and do not correspond to particular multipole moments Q_{lm}. The same rank-deficiency problems apply to the atom-centered PC grids. However, in that case, most of the eigenvectors do not have a direct correspondence to the multipole moments, which leads to even wider spread-out of the singular values/curvatures.

These insights can suggest several ways to alleviate the ill-conditioning of the problem. For instance, the buried atom problem can be addressed by truncating the sloppy singular vectors with dominant contribution from these atom, instead of introducing additional restraining functions[3,6,36,37] that can negatively affect the overall electrostatic properties of the molecule.[22,56] Also, slight deviations of the total charge of the fitted PC solution can be fixed by adjusting the stiff total-charge vector $\mathbf{u}_1$ and the corresponding coordinate Q_0^{mol}, rather than introducing a Lagrange multiplier that increases the rank-deficiency of the Hessian matrix.[21,22]

The results presented here can help further application of the PC model in biomolecular simulations. Although the force fields using point charges may not be as accurate as the force fields that explicitly include multipoles and/or polarization effects, the simplicity and computational efficiency of the PC model has ensured its continued survival.[18] In fact, representation of multipoles using the Lebedev grid PC model can provide an alternative to the multipole moment expansion;[57] it also can be used to extend recently proposed distributed charge model.[19,20]

VIII. COMPUTATIONAL DETAILS

MEP and multipole moments were calculated at the B3LYP/aug-cc-pVDZ level[58–61] as implemented in Q-Chem package.[62] For atom-centered PC fitting, the reference MEP was generated as the cubic grid with linear density 2.8 points/Å, followed by the removal of the points outside of 1.0-2.0 van der Waals radii range around each atom (vdW grid). For the two-sphere PC model, the Lebedev quadrature rules were used as implemented in PyQuante package.[63,64] Charge fitting procedures were implemented in the in-house developed *fftoolbox* Python library.[65] SVD was performed using *numpy* library.[66] Spherical harmonics were accessed from *scipy* library.[67]

ACKNOWLEDGMENTS

[1]S. R. Cox and D. E. Williams, J. Comput. Chem. **2**, 304 (1981).

[2]C. Singh and P. A. Kollman, J. Comput. Chem. **5**, 129 (1984).

[3]C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).

[4]K. Hinsen and B. Roux, J. Comput. Chem. **18**, 368 (1997).

[5]A. C. Simmonett, A. T. B. Gilbert, and P. M. W. Gill, Mol. Phys. **103**, 2789 (2005).

[6]F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, Phys. Chem. Chem. Phys. **12**, 7821 (2010).

[7]L. Huang and B. Roux, J. Chem. Theory Comput. **9**, 3543 (2013).

[8]A. W. Götz, D. Bucher, S. Lindert, and J. A. McCammon, J. Chem. Theory Comput. **10**, 1631 (2014).

[9]C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, and R. C. Walker, J. Chem. Theory Comput. **10**, 865 (2014).

[10]Y. A. Arnautova, R. Abagyan, and M. Totrov, J. Chem. Theory Comput. **11**, 2167 (2015).

[11]K. G. Sprenger, V. Jaeger, and J. Pfaendtner, J. Phys. Chem. B **119**, 5882 (2015).

[12]A. Mukhopadhyay, I. S. Tolokh, and A. V. Onufriev, J. Phys. Chem. B **119**, 6092 (2015).

[13]J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11**, 3696 (2015).

[14]A. Gabrieli, M. Sant, P. Demontis, and G. B. Suffritti, J. Chem. Theory Comput. **11**, 3829 (2015).

[15]S. Cardamone, T. J. Hughes, and P. L. A. Popelier, Phys. Chem. Chem. Phys. **16**, 10367 (2014).

[16]C. Kramer, A. Spinn, and K. R. Liedl, J. Chem. Theory Comput. **10**, 4488 (2014).

[17]P. G. Karamertzanis and C. C. Pantelides, Mol. Simul. **30**, 413 (2004).

[18]D. S. Cerutti, J. E. Rice, W. C. Swope, and D. A. Case, J. Phys. Chem. B **117**, 2328 (2013).

[19]Q. Gao, S. Yokojima, D. G. Fedorov, K. Kitaura, M. Sakurai, and S. Nakamura, Chem. Phys. Lett. **593**, 165 (2014).

[20]M. Devereux, S. Raghunathan, D. G. Fedorov, and M. Meuwly, J. Chem. Theory Comput. **10**, 4229 (2014).

[21]M. M. Francl, C. Carey, and D. M. Gange, J. Comput. Chem. **17**, 367 (1995).

[22]E. Sigfridsson, U. L. F. Ryde, T. Chemistry, C. Centre, and P. O. Box, J. Comput. Chem. **19**, 377 (1997).

[23]S. Jakobsen and F. Jensen, J. Chem. Theory Comput. **10**, 5493 (2014).

[24]S. M. Tschampel, M. R. Kennerty, and R. J. Woods, J. Chem. Theory Comput. **3**, 1721 (2007).

[25]R. W. Dixon and P. A. Kollman, J. Comput. Chem. **18**, 1632 (1997).

[26]T. R. Stouch and D. E. Williams, J. Comput. Chem. **13**, 622 (1992).

[27]T. R. Stouch and D. E. Williams, J. Comput. Chem. **14**, 858 (1993).

[28]C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).

[29]R. J. Woods, M. Khalil, W. Pell, S. H. Moffat, and V. H. Smith, J. Comput. Chem. **11**, 297 (1990).

[30]C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Series in Automatic Computation (Prentice-Hall, Englewood Cliffs, NJ 07632, USA, 1974).

[31]P. C. Hansen, *Discrete Inverse Problems - Insight and Algorithms* (SIAM, Philadelphia, USA, 2010).

[32]B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Science **342**, 604 (2013).

[33]M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, J. Chem. Phys. **143**, 010901 (2015).

[34]B. H. Besler, K. M. Merz, and P. A. Kollman, J. Comput. Chem. **11**, 431 (1990).

[35]P. Cieplak, W. D. Cornell, C. Bayly, and P. A. Kollman, J. Comput. Chem. **16**, 1357 (1995).

[36]J. Zeng, L. Duan, J. Z. H. Zhang, and Y. Mei, J. Comput. Chem. **34**, 847 (2013).

[37]S. K. Burger, J. Schofield, and P. W. Ayers, J. Phys. Chem. B **117**, 14960 (2013).

[38]M. V. Ivanov, M. R. Talipov, and Q. K. Timerghazin, J. Phys. Chem. A **119**, 1422 (2015).

[39]P. C. Hansen, V. Pereyra, and G. Scherer, *Least Squares Data Fitting* (The Johns Hopkins University Press, Baltimore, MD, USA, 2013).

[40]A. Laio, J. VandeVondele, and U. Rothlisberger, J. Phys. Chem. B **106**, 7300 (2002).

[41]C. W. Groetsch, J. Phys.: Conf. Ser. **73**, 012001 (2007).

[42]S. Roman, *Advanced Linear Algebra*, 3rd ed. (Springer, New York, NY, USA, 2008).

[43]K. S. Brown and J. P. Sethna, Phys. Rev. E **68**, 021904 (2003).

[44]A. Stone, *Theory of Intermolecular Forces* (Oxford University Press Inc., New York, NY, USA, 1996).

[45]J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (John Wiley & Sons, Inc., 1999).

[46]For practical purpose, we use real-valued spherical harmonics, thus to compact the derivation and not to obscure the main idea of the work Y_{lm} with, $m < 0$ ($m > 0$) corresponds to Y_{lms} (Y_{lmc}) in standard notation.

[47]S. S. Leang, F. Zahariev, and M. S. Gordon, J. Chem. Phys. **136**, 104101 (2012).

[48]S. N. Steinmann, C. Piemontesi, A. Delachat, and C. Corminboeuf, J. Chem. Theory Comput. **8**, 1629 (2012).

[49]R. M. Parrish, E. G. Hohenstein, T. J. Martínez, and C. D. Sherrill, J. Chem. Phys. **138**, 194107 (2013).

[50]V. I. Lebedev and D. N. Laikov, Doklady Akademii Nauk **366**, 741 (1999).

[51]See supplementary material at http://dx.doi.org/10.1063/1.4932105 for the Lebedev grid properties, SVD, and total charge constraint results.

[52]C. Ahrens and G. Beylkin, Proc. R. Soc. A **465**, 3103 (2009).

[53]V. I. Lebedev, USSR Comput. Math. Math. **16**, 10 (1976).

[54]I. Sloan, J. Approximation Theory **83**, 238 (1995).

[55]The slight differences in the resulting PC and multipole values obtained with the two methods arise due to the finite radius R of the probe sphere S_R used in the numerical approach. As R increases, the probe sphere S_R entirely encompasses the molecular charge density, and the multipole moments of the charged sphere Q_{lm}^{Sa} converge to the true molecular multipole moments Q_{lm}^{mol} (Fig. 7 in supplementary material).

[56]E. Vohringer-Martinez, T. Verstraelen, and P. W. Ayers, J. Phys. Chem. B **118**, 9871 (2014).

[57]D. M. Rogers, J. Chem. Phys. **142**, 074101 (2015).

[58]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[59]A. D. Becke, J. Phys. Chem. **98**, 5648 (1993).

[60]P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).

[61]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[62]Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).

[63]See https://github.com/gabrielelanaro/pyquante for PyQuante: Python Quantum Chemistry.

[64]See https://github.com/rpmuller/pyquante2 for PyQuante2: Python Quantum Chemistry.

[65]See https://github.com/maxivanoff/fftoolbox for Force field tool box: Python library to optimize force fields for molecular mechanics.

[66]S. van der Walt, S. Colbert, and G. Varoquaux, Comput. Sci. Eng. **13**, 22 (2011).

[67]See http://www.scipy.org/ for SciPy: Open source scientific tools for Python.