ORIGINAL PAPER

An improved parameterization procedure for NDDO-descendant semi-empirical methods

Adrian Wee Wen Ong[1,2] · Steve Yueran Cao[1,2] · Leong Chuan Kwek[2,3,4,5]

Abstract

Concept MNDO-based semi-empirical methods in quantum chemistry have found widespread application in the modelling of large and complex systems. A method for the analytic evaluation of first and second derivatives of molecular properties against semi-empirical parameters in MNDO-based NDDO-descendant models is presented, and the resultant parameter Hessian is compared against the approximant currently used in parameterization for the PMx models.

Methods As a proof of concept, the exact parameter Hessian is employed in a limited reparameterization of MNDO for the elements C, H, N, O and F using 1206 molecules for reference data (heats of formation, ionization energies, dipole moments and reference geometries). The correctness of our MNDO implementation was verified by comparing the calculated molecular properties with the MOPAC program.

Keywords NDDO · MNDO · Quantum chemistry

Introduction

In modern computational chemistry, semi-empirical methods based on the neglect of diatomic differential overlap (NDDO) [1–12] have found widespread applications in studies where more computationally intensive ab initio methods are unfeasible. While semi-empirical model Hamiltonians developed via machine-learning have gained significant attention [13, 14], recent work on the use of machine-learning techniques to construct correction terms for molecular properties calculated at lower levels of theory [15, 16] or develop semi-empirical molecular Hamiltonians [17] suggests that semi-empirical methods can achieve chemical accuracy. Essential to the

success of a developed semi-empirical model is a robust parameterization procedure that allows the model to best reproduce experimental data; as such, parameter optimization is an important area of study for the development of effective and accurate semi-empirical models.

While alternative formulations based on the NDDO approximation [18–22] have been proposed in recent years, most NDDO-descendant semi-empirical models [1, 2, 5, 9] use an identical formalism for the construction of the one-electron matrix $\mathbf{H}$ and two-electron matrix $\mathbf{G}$; with the exception of minor corrections to the asymptotic behaviour of two-electron integrals in PM7 [10], these methods also employ the same approximations and semi-empirical expressions [23, 24] for the evaluation of relevant molecular integrals. The difference between most NDDO-descendant models hence lies only in the parameter values chosen as well as the empirical expressions for core-core repulsion terms, which have been modified significantly between models [2, 9–11]; as such, any parameterization scheme developed for one NDDO-descendant model can be readily applied with few modifications to other NDDO-descendant models.

As noted by Stewart when developing PM7 [10], the parameter Hessian can be used to determine the nature of stationary points detected during parameterization; likewise, accurate second-derivative information can be used to great effect in searching for minima on surfaces. Construction of

✉ Adrian Wee Wen Ong
h1710105@nushigh.edu.sg

1 NUS High School of Mathematics and Science, 20 Clementi Avenue 1, 129957 Singapore, Singapore

2 Centre for Quantum Technologies, National University of Singapore, 117543 Singapore, Singapore

3 MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, 117543 Singapore, Singapore

4 National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, 637616 Singapore, Singapore

5 School of Electrical and Electronic Engineering Block S2.1, 50 Nanyang Avenue, 639798 Singapore, Singapore

the parameter Hessian via direct differentiation of the error function appears straightforward; in differentiating the error function $\mathcal{S}$ with respect to parameters $^{Z_A}p_i$ and $^{Z_B}p_i$, we find that

$$\mathcal{S} = \sum_{\alpha} \mathcal{C}_{\alpha}^2 \left(\xi_{\alpha}^{\text{ref}} - \xi_{\alpha} \right)^2 \tag{1}$$

$$\frac{\partial^2 \mathcal{S}}{\partial^{Z_A}p_i \partial^{Z_B}p_j} = 2 \sum_{\alpha} \mathcal{C}_{\alpha}^2 \left[\left(\xi_{\alpha} - \xi_{\alpha}^{\text{ref}} \right) \frac{\partial^2 \xi_{\alpha}}{\partial^{Z_A}p_i \partial^{Z_B}p_j} + \frac{\partial \xi_{\alpha}}{\partial^{Z_A}p_i} \frac{\partial \xi_{\alpha}}{\partial^{Z_B}p_j} \right] \tag{2}$$

It should be noted, however, that the parameter Hessian constructed in the development of PM7 appears to neglect second derivatives of the reference functions; as given in [10], the expression for $\frac{\partial^2 \mathcal{S}}{\partial^{Z_A}p_i \partial^{Z_B}p_j}$ appears to be

$$\frac{\partial^2 \mathcal{S}^{\text{PM7}}}{\partial^{Z_A}p_i \partial^{Z_B}p_j} = 2 \sum_{\alpha} \mathcal{C}_{\alpha}^2 \frac{\partial \xi_{\alpha}}{\partial^{Z_A}p_i} \frac{\partial \xi_{\alpha}}{\partial^{Z_B}p_j} \tag{3}$$

The neglect of the second-derivative term is expected to significantly affect the nature of the constructed Hessian matrix and its eigenvalues, impacting the quality of parameter optimization.

Evaluation of parameter derivatives of molecular properties via finite difference, as in standard procedure, results in numerical instability and may lead to an undesirable irreproducibility in results; thus, an analytical method for evaluation of parameter first and second derivatives is sought. This requires the evaluation of second derivatives of the density matrix as the idempotency condition [25] does not apply for derivatives of the ionization energy or dipole moment; an efficient method for solution of the second order coupled-perturbed Hartree–Fock (CPHF) equations is hence necessary for analytical derivative evaluation. NDDO methods formally operate in the Lowdin basis, where the overlap matrix between basis functions is substituted for the identity matrix; thus, the relevant equations for CPHF equation solution are greatly simplified and may be easily implemented.

The second-order CPHF equations under the NDDO approximation

The form of the first-order CPHF equations in the Lowdin basis for both closed- and open-shell species under the NDDO approximation has been well-documented [26] due to their implementation in geometry optimization routines [26, 27]; the second-order CPHF equations have been presented only in the MO basis [28] and are hence given in analogous form for the UHF case in the Lowdin basis.

All equations presented in this section are based on the Unrestricted Hartree–Fock (UHF) formalism; equations for

the restricted case are obtained as a special case of these results. Matrices for a specific spin (e.g. the Fock matrices for alpha- and beta-spin) are denoted as $^{\sigma}\mathbf{M}$, where $\sigma \in \{\alpha, \beta\}$ is an arbitrary spin. $\mathbf{P} = {}^{\alpha}\mathbf{P} + {}^{\beta}\mathbf{P}$ represents the density matrix in the Lowdin basis, which is the sum of the alpha- and beta-spin density matrices.

We first define the generalized Coulomb and exchange matrices $\mathbf{J}(\boldsymbol{\Delta})$ and $\mathbf{K}(\boldsymbol{\Lambda})$ as well as their associated static derivatives as the contraction of the relevant 2-electron integrals with arbitrary matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Lambda}$:

$$J_{\mu\nu}(\boldsymbol{\Delta}) = \begin{cases} \sum_{\lambda \in A} \Delta_{\lambda\lambda}(\mu\nu|\lambda\lambda) + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma}(\mu\nu|\lambda\sigma), \mu = \nu, \mu, \nu \in A \\ 2\Delta_{\mu\nu}(\mu\nu|\mu\nu) + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma}(\mu\nu|\lambda\sigma), \mu, \nu \in A \\ 0, \mu \in A, \nu \in B \neq A \end{cases} \tag{4}$$

$$J_{\mu\nu}^{q_1}(\boldsymbol{\Delta}) = \begin{cases} \sum_{\lambda \in A} \Delta_{\lambda\lambda} \frac{d(\mu\nu|\lambda\lambda)}{dq_1} + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma} \frac{d(\mu\nu|\lambda\sigma)}{dq_1}, \mu = \nu, \mu, \nu \in A \\ 2\Delta_{\mu\nu} \frac{d(\mu\nu|\mu\nu)}{dq_1} + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma} \frac{d(\mu\nu|\lambda\sigma)}{dq_1}, \mu, \nu \in A \\ 0, \mu \in A, \nu \in B \neq A \end{cases} \tag{5}$$

$$J_{\mu\nu}^{q_1 q_2}(\boldsymbol{\Delta}) = \begin{cases} \sum_{\lambda \in A} \Delta_{\lambda\lambda} \frac{d^2(\mu\nu|\lambda\lambda)}{dq_1 dq_2} + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma} \frac{d^2(\mu\nu|\lambda\sigma)}{dq_1 dq_2}, \mu = \nu, \mu, \nu \in A \\ 2\Delta_{\mu\nu} \frac{d^2(\mu\nu|\mu\nu)}{dq_1 dq_2} + \sum_{\lambda,\sigma \in B \neq A} \Delta_{\lambda\sigma} \frac{d^2(\mu\nu|\lambda\sigma)}{dq_1 dq_2}, \mu, \nu \in A \\ 0, \mu \in A, \nu \in B \neq A \end{cases} \tag{6}$$

$$K_{\mu\nu}(\boldsymbol{\Lambda}) = \begin{cases} \sum_{\lambda \in A} \Lambda_{\lambda\lambda}(\mu\lambda|\nu\lambda), \mu = \nu, \mu, \nu \in A \\ \Lambda_{\mu\nu}[(\mu\nu|\mu\nu) + (\mu\mu|\nu\nu)], \mu, \nu \in A \\ \sum_{\substack{\lambda \in A \\ \sigma \in B}} \Lambda_{\lambda\sigma}(\mu\lambda|\nu\sigma), \mu \in A, \nu \in B \neq A \end{cases} \tag{7}$$

$$K_{\mu\nu}^{q_1}(\boldsymbol{\Lambda}) = \begin{cases} \sum_{\lambda \in A} \Lambda_{\lambda\lambda} \frac{d(\mu\lambda|\nu\lambda)}{dq_1}, \mu = \nu, \mu, \nu \in A \\ \Lambda_{\mu\nu}\left[\frac{d(\mu\nu|\mu\nu)}{dq_1} + \frac{d(\mu\mu|\nu\nu)}{dq_1}\right], \mu, \nu \in A \\ \sum_{\substack{\lambda \in A \\ \sigma \in B}} \Lambda_{\lambda\sigma} \frac{d(\mu\lambda|\nu\sigma)}{dq_1}, \mu \in A, \nu \in B \neq A \end{cases} \tag{8}$$

$$K_{\mu\nu}^{q_1 q_2}(\boldsymbol{\Lambda}) = \begin{cases} \sum_{\lambda \in A} \Lambda_{\lambda\lambda} \frac{d^2(\mu\lambda|\nu\lambda)}{dq_1 dq_2}, \mu = \nu, \mu, \nu \in A \\ \Lambda_{\mu\nu}\left[\frac{d^2(\mu\nu|\mu\nu)}{dq_1 dq_2} + \frac{d^2(\mu\mu|\nu\nu)}{dq_1 dq_2}\right], \mu, \nu \in A \\ \sum_{\substack{\lambda \in A \\ \sigma \in B}} \Lambda_{\lambda\sigma} \frac{d^2(\mu\lambda|\nu\sigma)}{dq_1 dq_2}, \mu \in A, \nu \in B \neq A \end{cases} \tag{9}$$

In conventional NDDO methods [1–11], the core matrix is given by

$$\mathbf{H}_{\mu\nu} = \begin{cases} {}^{Z_A}U_{\mu\mu} + \sum_{B \neq A} V_{\mu\mu,B}, \mu = \nu, \mu, \nu \in A \\ \sum_{B \neq A} V_{\mu\nu,B}, \mu, \nu \in A \\ \beta_{\mu\nu}, \mu \in A, \nu \in B \neq A \end{cases} \tag{10}$$

In the above, $V_{\mu\nu,B}$ represents the two-centre nuclear-electron attraction integral between basis functions μ, ν and atom B; $\beta_{\mu\nu} = \frac{\beta_\mu + \beta_\nu}{2} S_{\mu\nu}$ represents the resonance integral between two basis functions on different atoms.

The Fock matrix in the NDDO formalism can hence be given as follows:

$$^\sigma\mathbf{F} = \mathbf{H} + \mathbf{J}(\mathbf{P}) - \mathbf{K}(^\sigma\mathbf{P}) \tag{11}$$

In matrix form, the first-order orbital coefficients $^\sigma\mathbf{x}^{q_1}$ are related to the direct derivative of the (AO) coefficient matrix $^\sigma\mathbf{C}$ as follows:

$$\frac{d^\sigma\mathbf{C}}{dq_1} = {}^\sigma\mathbf{C}\,^\sigma\mathbf{x}^{q_1} \tag{12}$$

Analogously, the second-order orbital coefficients can be obtained via differentiation of the first-order orbital coefficients:

$$\frac{d^\sigma\mathbf{x}^{q_1}}{dq_2} = {}^\sigma\boldsymbol{\gamma}^{q_1 q_2} - {}^\sigma\mathbf{x}^{q_2}\,^\sigma\mathbf{x}^{q_1} \tag{13}$$

Application of the orthonormality condition $^\sigma\mathbf{C}^{\mathrm{T}\,\sigma}\mathbf{C} = 1$ yields the relevant commutator relations for the first- and second-order orbital coefficients:

$$^\sigma x^{q_1}_{ij} + {}^\sigma x^{q_1}_{ji} = 0 \tag{14}$$

$$^\sigma\gamma^{q_1 q_2}_{ij} + {}^\sigma\gamma^{q_1 q_2}_{ji} = \sum_k \left({}^\sigma x^{q_1}_{ik}\,^\sigma x^{q_2}_{kj} + {}^\sigma x^{q_2}_{ik}\,^\sigma x^{q_1}_{kj} \right) \tag{15}$$

The derivatives of the off-diagonal elements of the MO-basis Fock matrix must be zero due to the variational condition:

$$^\sigma\mathfrak{F} = {}^\sigma\mathbf{C}^{\mathrm{T}\,\sigma}\mathbf{F}\,^\sigma\mathbf{C} \tag{16}$$

$$\left(\frac{d^\sigma\mathfrak{F}}{dq_1} \right)_{ij} = 0, \quad \left(\frac{d^{2\sigma}\mathfrak{F}}{dq_1 dq_2} \right)_{ij} = 0, \quad i \neq j \tag{17}$$

The first-order CPHF equations are well-documented in the literature and are hence only presented for completeness.

Evaluation of the first derivative $\frac{d^\sigma\mathfrak{F}}{dq_1}$ yields the following:

$$\frac{d^\sigma\mathfrak{F}}{dq_1} = ({}^\sigma\mathbf{x}^{q_1})^{\mathrm{T}\,\sigma}\mathfrak{F} + {}^\sigma\mathbf{C}^{\mathrm{T}} \frac{d^\sigma\mathbf{F}}{dq_1}\,^\sigma\mathbf{C} + {}^\sigma\mathfrak{F}\,^\sigma\mathbf{x}^{q_1} \tag{18}$$

$$\frac{d^\sigma\mathbf{F}}{dq_1} = \frac{d\mathbf{H}}{dq_1} + \mathbf{J}^{q_1}(\mathbf{P}) - \mathbf{K}^{q_1}(^\sigma\mathbf{P}) + \mathbf{J}\left(\frac{d\mathbf{P}}{dq_1} \right) - \mathbf{K}\left(\frac{d^\sigma\mathbf{P}}{dq_1} \right) \tag{19}$$

The "static" derivative term $\frac{d\mathbf{H}}{dq_1} + \mathbf{J}^{q_1}(\mathbf{P}) - \mathbf{K}^{q_1}(^\sigma\mathbf{P})$ is referred to as $^\sigma\mathbf{F}^{q_1}$ and the response term $\mathbf{J}\left(\frac{d\mathbf{P}}{dq_1} \right) - \mathbf{K}\left(\frac{d^\sigma\mathbf{P}}{dq_1} \right)$ is referred to as $^\sigma\mathbf{R}^{q_1}$; collection of terms yields

$$^\sigma\mathfrak{F}^{q_1} = {}^\sigma\mathbf{C}^{\mathrm{T}} \left(\frac{d\mathbf{H}}{dq_1} + \mathbf{J}^{q_1}(\mathbf{P}) - \mathbf{K}^{q_1}(^\sigma\mathbf{P}) \right) {}^\sigma\mathbf{C} \tag{20}$$

$$^\sigma\mathfrak{R}^{q_1} = {}^\sigma\mathbf{C}^{\mathrm{T}} \left(\mathbf{J}\left(\frac{d\mathbf{P}}{dq_1} \right) - \mathbf{K}\left(\frac{d^\sigma\mathbf{P}}{dq_1} \right) \right) {}^\sigma\mathbf{C} \tag{21}$$

$$\begin{aligned} \frac{d^\sigma\mathfrak{F}}{dq_1} &= ({}^\sigma\mathbf{x}^{q_1})^{\mathrm{T}\,\sigma}\mathfrak{F} + {}^\sigma\mathbf{C}^{\mathrm{T}}({}^\sigma\mathbf{F}^{q_1} + {}^\sigma\mathbf{R}^{q_1}){}^\sigma\mathbf{C} + {}^\sigma\mathfrak{F}\,^\sigma\mathbf{x}^{q_1} \\ &= ({}^\sigma\mathbf{x}^{q_1})^{\mathrm{T}\,\sigma}\mathfrak{F} + {}^\sigma\mathfrak{F}^{q_1} + {}^\sigma\mathfrak{R}^{q_1} + {}^\sigma\mathfrak{F}\,^\sigma\mathbf{x}^{q_1} \end{aligned} \tag{22}$$

Thus, the first-order CPHF equations are given as follows, as reported in [26, 27]:

$$\left(\frac{d^\sigma\mathfrak{F}}{dq_1} \right)_{ij} = {}^\sigma x^{q_1}_{ji}\,^\sigma\epsilon_j + {}^\sigma x^{q_1}_{ij}\,^\sigma\epsilon_i + {}^\sigma\mathfrak{F}^{q_1}_{ij} + {}^\sigma\mathfrak{R}^{q_1}_{ij} = 0, \quad i \neq j \tag{23}$$

$$\left({}^\sigma\epsilon_j - {}^\sigma\epsilon_i \right){}^\sigma x^{q_1}_{ij} - {}^\sigma\mathfrak{R}^{q_1}_{ij} = {}^\sigma\mathfrak{F}^{q_1}_{ij} \tag{24}$$

To evaluate $^\sigma\mathfrak{R}^{q_1}_{ij}$, $\frac{d^\sigma\mathbf{P}}{dq_1}$ must be cast in terms of $^\sigma\mathbf{x}^{q_1}$; since $^\sigma\mathbf{P} = {}^\sigma\mathbf{C}^{\mathrm{T}}_{occ}\,^\sigma\mathbf{C}_{occ}$,

$$\left(\frac{d^\sigma\mathbf{P}}{dq_1} \right)_{\mu\nu} = \sum_{i \in occ.} \sum_j \left({}^\sigma c_{\mu j}\,^\sigma c_{\nu i} + {}^\sigma c_{\mu i}\,^\sigma c_{\nu j} \right) x^{q_1}_{ji} = -\sum_{i \in occ.} \sum_{j \in virt.} \left({}^\sigma c_{\mu j}\,^\sigma c_{\nu i} + {}^\sigma c_{\mu i}\,^\sigma c_{\nu j} \right) x^{q_1}_{ij} \tag{25}$$

Lastly, the expression for $\frac{d\mathbf{H}}{dq_1}$ is of Eq. (10):

$$\left(\frac{d\mathbf{H}}{dq_1} \right)_{\mu\nu} = \begin{cases} \dfrac{d^{Z_A} U_{\mu\mu}}{dq_1} + \sum_{B \neq A} \dfrac{dV_{\mu\mu,B}}{dq_1}, & \mu = \nu, \ \mu, \nu \in A \\[2ex] \sum_{B \neq A} \dfrac{dV_{\mu\nu,B}}{dq_1}, & \mu, \nu \in A \\[2ex] \dfrac{d\beta_{\mu\nu}}{dq_1}, & \mu \in A, \ \nu \in B \neq A \end{cases} \tag{26}$$

The first-order CPHF equations are hence linear in the first-order coefficients $x^{q_1}_{ij}$.

Since only the occupied-virtual block of $^\sigma\mathbf{x}^{q_1}$ is necessary to solve for $\frac{d^\sigma\mathbf{P}}{dq_1}$, the remaining elements of $^\sigma\mathbf{x}^{q_1}$ are evaluated afterwards via direct substitution:

$$^\sigma x^{q_1}_{ij} = \begin{cases} 0, & i = j \\[2ex] \dfrac{{}^\sigma\mathfrak{F}^{q_1} + {}^\sigma\mathfrak{R}^{q_1}}{{}^\sigma\epsilon_j - {}^\sigma\epsilon_i}, & i \neq j \end{cases} \tag{27}$$

The full matrix $^\sigma\mathbf{x}^{q_1}$, not just the occupied-virtual block, is necessary for solution of the second-order CPHF equations. These are obtained via further differentiation of Eq. (18):

$$\begin{aligned} \frac{d^{2\sigma}\mathfrak{F}}{dq_1 dq_2} &= \left(\frac{d^\sigma\mathbf{x}^{q_1}}{dq_2} \right)^{\mathrm{T}\,\sigma}\mathfrak{F} + ({}^\sigma\mathbf{x}^{q_1})^{\mathrm{T}} \frac{d^\sigma\mathfrak{F}}{dq_2} + ({}^\sigma\mathbf{x}^{q_2})^{\mathrm{T}} \left({}^\sigma\mathfrak{F}^{q_1} + {}^\sigma\mathfrak{R}^{q_1} \right) \\ &\quad + {}^\sigma\mathbf{C}^{\mathrm{T}} \frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2}\,^\sigma\mathbf{C} + \left({}^\sigma\mathfrak{F}^{q_1} + {}^\sigma\mathfrak{R}^{q_1} \right) {}^\sigma\mathbf{x}^{q_2} + \frac{d^\sigma\mathfrak{F}}{dq_2}\,^\sigma\mathbf{x}^{q_1} + {}^\sigma\mathfrak{F} \frac{d^\sigma\mathbf{x}^{q_1}}{dq_2} \end{aligned} \tag{28}$$

Simplification yields:

$$\left(\frac{d^{2\sigma}\mathfrak{F}}{dq_1 dq_2}\right)_{ij} = \left(^\sigma \mathbf{C}^T \frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2}{}^\sigma \mathbf{C}\right)_{ij} + {}^\sigma \gamma_{ji}^{q_1 q_2}{}^\sigma \epsilon_j$$
$$+ \left((^\sigma \mathbf{x}^{q_1})^T \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right) + (^\sigma \mathbf{x}^{q_1})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_2} + \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right)^\sigma \mathbf{x}^{q_1}\right)_{ij} + {}^\sigma \gamma_{ij}^{q_1 q_2}{}^\sigma \epsilon_i$$ (29)
$$+ \left((^\sigma \mathbf{x}^{q_2})^T \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_1}\right) + (^\sigma \mathbf{x}^{q_2})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_1} + \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_1}\right)^\sigma \mathbf{x}^{q_2}\right)_{ij}$$

Application of the commutator relation further reduces the expression complexity:

$$\left(\frac{d^{2\sigma}\mathfrak{F}}{dq_1 dq_2}\right)_{ij} = \left(^\sigma \mathbf{C}^T \frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2}{}^\sigma \mathbf{C}\right)_{ij} - (^\sigma \epsilon_j - {}^\sigma \epsilon_i)^\sigma \gamma_{ij}^{q_1 q_2} + ((^\sigma \mathbf{x}^{q_1\sigma}\mathbf{x}^{q_2} + {}^\sigma \mathbf{x}^{q_2\sigma}\mathbf{x}^{q_1})^\sigma \boldsymbol{\epsilon})_{ij}$$
$$+ \left((^\sigma \mathbf{x}^{q_1})^T \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right) + (^\sigma \mathbf{x}^{q_1})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_2} + \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right)^\sigma \mathbf{x}^{q_1}\right)_{ij}$$ (30)
$$+ \left((^\sigma \mathbf{x}^{q_2})^T \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_2}\right) + (^\sigma \mathbf{x}^{q_2})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_1} + \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_1}\right)^\sigma \mathbf{x}^{q_2}\right)_{ij}$$

Evaluation of $\frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2}$ is performed analogously to Eq. (19), albeit with significantly more terms in the resultant expression:

$$\frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2} = \frac{d^2 \mathbf{H}}{dq_1 dq_2} + \mathbf{J}^{q_1 q_2}(\mathbf{P}) + \mathbf{J}^{q_1}\left(\frac{d\mathbf{P}}{dq_2}\right) + \mathbf{J}^{q_2}\left(\frac{d\mathbf{P}}{dq_1}\right) - \mathbf{K}^{q_1 q_2}(^\sigma \mathbf{P})$$
$$- \mathbf{K}^{q_1}\left(\frac{d^\sigma \mathbf{P}}{dq_2}\right) - \mathbf{K}^{q_2}\left(\frac{d^\sigma \mathbf{P}}{dq_1}\right) + \mathbf{J}\left(\frac{d^2 \mathbf{P}}{dq_1 dq_2}\right) - \mathbf{K}\left(\frac{d^2 \mathbf{P}}{dq_1 dq_2}\right)$$ (31)

Lastly, direct differentiation of the density matrix yields

$$\left(\frac{d^{2\sigma}\mathbf{P}}{dq_1 dq_2}\right)_{\mu\nu} = \sum_{i \in occ.} \sum_{j,k} \left(^\sigma c_{\mu k}{}^\sigma c_{\nu i} + {}^\sigma c_{\mu i}{}^\sigma c_{\nu k}\right)\left(^\sigma x_{kj}^{q_2\sigma} x_{ji}^{q_1} + {}^\sigma x_{kj}^{q_1\sigma} x_{ji}^{q_2}\right)$$
$$+ \sum_{i \in occ.} \sum_{j \in occ.} \sum_k {}^\sigma c_{\mu i}{}^\sigma c_{\nu j}\left(^\sigma x_{ik}^{q_1\sigma} x_{kj}^{q_2} + {}^\sigma x_{ik}^{q_2\sigma} x_{kj}^{q_1}\right)$$
$$+ \sum_{i \in occ.} \sum_{j,k} \left(^\sigma c_{\mu j}{}^\sigma c_{\nu k} + {}^\sigma c_{\mu k}{}^\sigma c_{\nu j}\right)^\sigma x_{ki}^{q_1\sigma} x_{ji}^{q_2}$$
$$- \sum_{i \in occ.} \sum_{j \in virt.} \left(^\sigma c_{\mu j}{}^\sigma c_{\nu i} + {}^\sigma c_{\mu i}{}^\sigma c_{\nu j}\right) \gamma_{ij}^{q_1 q_2}$$ (32)

Since the first three terms in $\left(\frac{d^{2\sigma}\mathbf{P}}{dq_1 dq_2}\right)_{\mu\nu}$ are independent of the second-order coefficients, we define

$$^\sigma \boldsymbol{\rho}_{\mu\nu}^{q_1 q_2} = - \sum_{i \in occ.} \sum_{j \in virt.} \left(^\sigma c_{\mu j}{}^\sigma c_{\nu i} + {}^\sigma c_{\mu i}{}^\sigma c_{\nu j}\right) \gamma_{ij}^{q_1 q_2}$$ (33)

$$^\sigma \varsigma_{\mu\nu}^{q_1 q_2} = \sum_{i \in occ.} \sum_{j,k} \left(^\sigma c_{\mu k}{}^\sigma c_{\nu i} + {}^\sigma c_{\mu i}{}^\sigma c_{\nu k}\right)\left(^\sigma x_{kj}^{q_2\sigma} x_{ji}^{q_1} + {}^\sigma x_{kj}^{q_1\sigma} x_{ji}^{q_2}\right)$$
$$+ \sum_{i \in occ.} \sum_{j \in occ.} \sum_k {}^\sigma c_{\mu i}{}^\sigma c_{\nu j}\left(^\sigma x_{ik}^{q_1\sigma} x_{kj}^{q_2} + {}^\sigma x_{ik}^{q_2\sigma} x_{kj}^{q_1}\right)$$
$$+ \sum_{i \in occ.} \sum_{j,k} \left(^\sigma c_{\mu j}{}^\sigma c_{\nu k} + {}^\sigma c_{\mu k}{}^\sigma c_{\nu j}\right)^\sigma x_{ki}^{q_1\sigma} x_{ji}^{q_2}$$ (34)

$$\frac{d^{2\sigma}\mathbf{P}}{dq_1 dq_2} = {}^\sigma \boldsymbol{\rho}^{q_1 q_2} + {}^\sigma \varsigma^{q_1 q_2}$$ (35)

The static derivative term in $\frac{d^{2\sigma}\mathbf{F}}{dq_1 dq_2}$ is hence termed $^\sigma \mathbf{F}^{q_1 q_2}$, with the response term correspondingly referred to as $^\sigma \mathbf{R}^{q_1 q_2}$:

$$^\sigma \mathbf{F}^{q_1 q_2} = \frac{d^2 \mathbf{H}}{dq_1 dq_2} + \mathbf{J}^{q_1 q_2}(\mathbf{P}) + \mathbf{J}^{q_1}\left(\frac{d\mathbf{P}}{dq_2}\right) + \mathbf{J}^{q_2}\left(\frac{d\mathbf{P}}{dq_1}\right)$$
$$- \mathbf{K}^{q_1 q_2}(^\sigma \mathbf{P}) - \mathbf{K}^{q_1}\left(\frac{d^\sigma \mathbf{P}}{dq_2}\right) - \mathbf{K}^{q_2}\left(\frac{d^\sigma \mathbf{P}}{dq_1}\right) + \mathbf{J}(\varsigma^{q_1 q_2}) - \mathbf{K}(^\sigma \varsigma^{q_1 q_2})$$ (36)

$$^\sigma \mathbf{R}^{q_1 q_2} = \mathbf{J}(\boldsymbol{\rho}^{q_1 q_2}) - \mathbf{K}(^\sigma \boldsymbol{\rho}^{q_1 q_2})$$ (37)

Accordingly, defining $^\sigma \mathfrak{F}^{q_1 q_2}$ and $^\sigma \mathfrak{R}^{q_1 q_2}$ yields the second-order CPHF equations:

$$^\sigma \mathfrak{F}^{q_1 q_2} = {}^\sigma \mathbf{C}^{T\sigma} \mathbf{F}^{q_1 q_2\sigma} \mathbf{C} + (^\sigma \mathbf{x}^{q_1\sigma}\mathbf{x}^{q_2} + {}^\sigma \mathbf{x}^{q_2\sigma}\mathbf{x}^{q_1})^\sigma \boldsymbol{\epsilon}$$
$$+ (^\sigma \mathbf{x}^{q_1})^T \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right) + (^\sigma \mathbf{x}^{q_1})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_2} + \left(^\sigma \mathfrak{F}^{q_2} + {}^\sigma \mathfrak{R}^{q_2}\right)^\sigma \mathbf{x}^{q_1}$$
$$+ (^\sigma \mathbf{x}^{q_2})^T \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_2}\right) + (^\sigma \mathbf{x}^{q_2})^{T\sigma} \boldsymbol{\epsilon}^\sigma \mathbf{x}^{q_1} + \left(^\sigma \mathfrak{F}^{q_1} + {}^\sigma \mathfrak{R}^{q_1}\right)^\sigma \mathbf{x}^{q_2}$$ (38)

$$^\sigma \mathfrak{R}^{q_1 q_2} = {}^\sigma \mathbf{C}^T (\mathbf{J}(\boldsymbol{\rho}^{q_1 q_2}) - \mathbf{K}(^\sigma \boldsymbol{\rho}^{q_1 q_2}))^\sigma \mathbf{C}$$ (39)

$$\left(\frac{d^2 \mathbf{H}}{dq_1 dq_2}\right)_{\mu\nu} = \begin{cases} \frac{d^2{}^{Z_A} U_{\mu\mu}}{dq_1 dq_2} + \sum_{B \neq A} \frac{d^2 V_{\mu\mu,B}}{dq_1 dq_2}, \mu = \nu, \mu, \nu \in A \\ \sum_{B \neq A} \frac{d^2 V_{\mu\nu,B}}{dq_1 dq_2}, \mu, \nu \in A \\ \frac{d^2 \beta_{\mu\nu}}{dq_1 dq_2}, \mu \in A, \nu \in B \neq A \end{cases}$$ (40)

$$(^\sigma \epsilon_j - {}^\sigma \epsilon_i)^\sigma \gamma_{ij}^{q_1 q_2} - {}^\sigma \mathfrak{R}_{ij}^{q_1 q_2} = {}^\sigma \mathfrak{F}_{ij}^{q_1 q_2}$$ (41)

The form of the second-order CPHF equations has intentionally been cast into a form that resemble the first-order CPHF equations; as such, the same algorithms [26, 29] employed to solve the first-order CPHF equations may be applied for solution of the second-order CPHF equations.

Derivatives of molecular properties under the MNDO formalism

The first derivatives $\frac{d(\Delta H_f)}{d^{Z_A}\alpha}$ and $\frac{d(\Delta H_f)}{d^{Z_A}E_{eisol}}$ are easily evaluated:

$$\frac{d(\Delta H_f)}{d^{Z_A}p} = k_{conv}\left(\frac{dE_{el}}{d^{Z_A}p} + \frac{dV_{core}}{d^{Z_A}p} - \sum_A \frac{d^{Z_A}E_{eisol}}{d^{Z_A}p}\right)$$ (42)

$$\frac{d(\Delta H_f)}{d^{Z_A}\alpha} = k_{conv}\sum_{C>B} \frac{dV_{BC}^{CRF}}{d^{Z_A}\alpha}, \frac{d(\Delta H_f)}{d^{Z_A}E_{eisol}} = -k_{conv} n_{Z_A}$$ (43)

If another NDDO-based semi-empirical method using the same formalism as MNDO for **H** and **G** [1–11] is required, the derivatives of ΔH_f with respect to the additional core-repulsion function parameters will depend on the expression for V_{BC}^{CRF} and can be easily obtained via direct differentiation.

The remaining derivatives (against $^{Z_A}\beta_s, {}^{Z_A}\beta_p, {}^{Z_A}U_{ss}, {}^{Z_A}U_{pp}, {}^{Z_A}\zeta_s, {}^{Z_A}\zeta_p$) are identical for [1–11] but will require modification for other methods (e.g. MNDO-F, OMx):

$$\frac{d(E_{el})}{d^{Z_A}p} = k_{conv}\left[\text{Tr}\left(\frac{d\mathbf{H}}{d^{Z_A}p}\mathbf{P}\right) + \frac{1}{2}\text{Tr}\left(\mathbf{J}^{Z_A p}(\mathbf{P})\mathbf{P} - \mathbf{K}^{Z_A p}(^\alpha \mathbf{P})^\alpha \mathbf{P} - \mathbf{K}^{Z_A p}(^\beta \mathbf{P})^\beta \mathbf{P}\right)\right]$$ (44)

Likewise,

$$\frac{\partial^2(\Delta H_f)}{\partial^{Z_A}\alpha \partial^{Z_B}\alpha} = k_{conv}\sum_{D>C}\frac{\partial^2 V_{CD}^{CRF}}{\partial^{Z_A}\alpha \partial^{Z_B}\alpha} \tag{45}$$

$$\frac{\partial^2(\Delta H_f)}{\partial^{Z_A}\alpha \partial^{Z_B}p_i} = 0 \forall^{Z_B}p_i \neq {}^{Z_B}\alpha \tag{46}$$

$$\frac{\partial^2(\Delta H_f)}{\partial^{Z_A}E_{eisol}\partial^{Z_B}p_i} = 0 \forall^{Z_B}p_i \tag{47}$$

$$\begin{aligned}\frac{\partial^2(\Delta H_f)}{\partial^{Z_A}p_i\partial^{Z_B}p_j} &= k_{conv}\mathrm{Tr}\left(\frac{d^2\mathbf{H}}{d^{Z_A}p_i d^{Z_B}p_j}\mathbf{P} + \frac{d\mathbf{H}}{d^{Z_A}p_i}\frac{d\mathbf{P}}{d^{Z_B}p_j}\right)\\&+\frac{1}{2}k_{conv}\mathrm{Tr}\left(\mathbf{J}^{Z_A p_i {}^{Z_B}p_j}(\mathbf{P})\mathbf{P} + \mathbf{J}^{Z_A p}\left(\frac{d\mathbf{P}}{d^{Z_B}p_j}\right)\mathbf{P} + \mathbf{J}^{Z_A p}(\mathbf{P})\frac{d\mathbf{P}}{d^{Z_B}p_j}\right)\\&-\frac{1}{2}k_{conv}\mathrm{Tr}\left(\mathbf{K}^{Z_A p_i {}^{Z_B}p_j}({}^\alpha\mathbf{P}){}^\alpha\mathbf{P} + \mathbf{K}^{Z_A p}\left(\frac{d^\alpha\mathbf{P}}{d^{Z_B}p_j}\right){}^\alpha\mathbf{P} + \mathbf{K}^{Z_A p}({}^\alpha\mathbf{P})\frac{d^\alpha\mathbf{P}}{d^{Z_B}p_j}\right)\\&-\frac{1}{2}k_{conv}\mathrm{Tr}\left(\mathbf{K}^{Z_A p_i {}^{Z_B}p_j}({}^\beta\mathbf{P}){}^\beta\mathbf{P} + \mathbf{K}^{Z_A p}\left(\frac{d^\beta\mathbf{P}}{d^{Z_B}p_j}\right){}^\beta\mathbf{P} + \mathbf{K}^{Z_A p}({}^\beta\mathbf{P})\frac{d^\beta\mathbf{P}}{d^{Z_B}p_j}\right)\end{aligned} \tag{48}$$

The expressions for $\frac{d\mathbf{H}}{d^{Z_A}p}$, $\frac{d^2\mathbf{H}}{d^{Z_A}p_i d^{Z_B}p_j}$, $\mathbf{J}^{Z_A p}(\boldsymbol{\Delta})$, $\mathbf{J}^{Z_A p_i {}^{Z_B}p_j}(\boldsymbol{\Delta})$, $\mathbf{K}^{Z_A p}(\boldsymbol{\Lambda})$ and $\mathbf{K}^{Z_A p_i {}^{Z_B}p_j}(\boldsymbol{\Lambda})$ for each of the parameters ${}^{Z_A}\beta_s, {}^{Z_A}\beta_p, {}^{Z_A}U_{ss}, {}^{Z_A}U_{pp}, {}^{Z_A}\zeta_s, {}^{Z_A}\zeta_p$ are detailed in the Supplementary Information.

Direct differentiation yields the first and second derivatives of the (semi-empirical) dipole moment and the energy matrix ${}^\sigma\boldsymbol{\epsilon}$ via which the ionization energy can be obtained:

$$\frac{d\langle\mu\rangle}{d^{Z_A}p} = \frac{\boldsymbol{\mu}\cdot\frac{d\boldsymbol{\mu}}{d^{Z_A}p}}{|\boldsymbol{\mu}|}, \tag{49}$$

$$\frac{d\mu_\tau}{d^{Z_A}p} = -2\sum_B\left(\frac{d^{Z_B}D_1}{d^{Z_A}p}P_{sp_\tau} + {}^{Z_B}D_1\frac{dP_{sp_\tau}}{d^{Z_A}p}\right) + \sum_B\tau_{CM,B}\left(Q_A - \sum_{m\in B}\frac{dP_{mm}}{d^{Z_A}p}\right) \tag{50}$$

$$\frac{d^2\langle\mu\rangle}{d^{Z_A}p_i d^{Z_B}p_j} = \frac{\boldsymbol{\mu}\cdot\frac{d^2\boldsymbol{\mu}}{d^{Z_A}\zeta_{x_1}d^{Z_B}\zeta_{x_2}} + \frac{d\boldsymbol{\mu}}{d^{Z_B}p_j}\cdot\frac{d\boldsymbol{\mu}}{d^{Z_A}p_i} - \frac{d\langle\mu\rangle}{d^{Z_A}p_i}\frac{d\langle\mu\rangle}{d^{Z_B}p_j}}{|\boldsymbol{\mu}|} \tag{51}$$

$$\begin{aligned}\frac{d^2\mu_\tau}{d^{Z_A}p_i d^{Z_B}p_j} &= \sum_B\tau_{CM,B}\left(Q_A - \sum_{m\in B}\frac{d^2P_{mm}}{d^{Z_A}p_i d^{Z_B}p_j}\right)\\&-2\sum_C\left(\frac{d^{Z_C}D_1}{d^{Z_A}p_i}\frac{dP_{sp_\tau}}{d^{Z_B}p_j} + \frac{d^{2Z_C}D_1}{d^{Z_A}p_i d^{Z_B}p_j}P_{sp_\tau} + \frac{d^{Z_C}D_1}{d^{Z_B}p_j}\frac{dP_{sp_\tau}}{d^{Z_A}p_i} + {}^{Z_C}D_1\frac{d^2P_{sp_\tau}}{d^{Z_A}p_i d^{Z_B}p_j}\right)\end{aligned} \tag{52}$$

$$\frac{d^\sigma\boldsymbol{\epsilon}}{d^{Z_A}p} = \frac{d^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_A}p}{}^\sigma\mathbf{F}^\sigma\mathbf{C} + {}^\sigma\mathbf{C}^{\mathrm{T}}\frac{d^\sigma\mathbf{F}}{d^{Z_A}p}{}^\sigma\mathbf{C} + {}^\sigma\mathbf{C}^{\mathrm{T}}{}^\sigma\mathbf{F}\frac{\partial^\sigma\mathbf{C}}{\partial^{Z_A}p} \tag{53}$$

$$\begin{aligned}\frac{d^2{}^\sigma\boldsymbol{\epsilon}}{d^{Z_A}p_i d^{Z_B}p_j} &= \frac{d^2{}^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_A}p_i d^{Z_B}p_j}{}^\sigma\mathbf{F}^\sigma\mathbf{C} + \frac{d^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_A}p_i}\frac{d^\sigma\mathbf{F}}{d^{Z_B}p_j}{}^\sigma\mathbf{C} + \frac{d^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_A}p_i}{}^\sigma\mathbf{F}\frac{d^\sigma\mathbf{C}}{d^{Z_B}p_j}\\&+\frac{d^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_B}p_j}\frac{d^\sigma\mathbf{F}}{d^{Z_A}p_i}{}^\sigma\mathbf{C} + {}^\sigma\mathbf{C}^{\mathrm{T}}\frac{d^\sigma\mathbf{F}}{d^{Z_A}p_i d^{Z_B}p_j}{}^\sigma\mathbf{C} + {}^\sigma\mathbf{C}^{\mathrm{T}}\frac{d^\sigma\mathbf{F}}{d^{Z_A}p_i}\frac{d^\sigma\mathbf{C}}{d^{Z_B}p_j}\\&+\frac{d^\sigma\mathbf{C}^{\mathrm{T}}}{d^{Z_B}p_j}{}^\sigma\mathbf{F}\frac{d^\sigma\mathbf{C}}{d^{Z_A}p_i} + {}^\sigma\mathbf{C}^{\mathrm{T}}\frac{d^\sigma\mathbf{F}}{d^{Z_B}p_j}\frac{d^\sigma\mathbf{C}}{d^{Z_A}p_i} + {}^\sigma\mathbf{C}^{\mathrm{T}}{}^\sigma\mathbf{F}\frac{d^2{}^\sigma\mathbf{C}}{d^{Z_A}p_i d^{Z_B}p_j}\end{aligned} \tag{54}$$

Derivatives of dipole moments are presented for the Restricted Hartree–Fock case, as no extension to UHF systems is necessary.
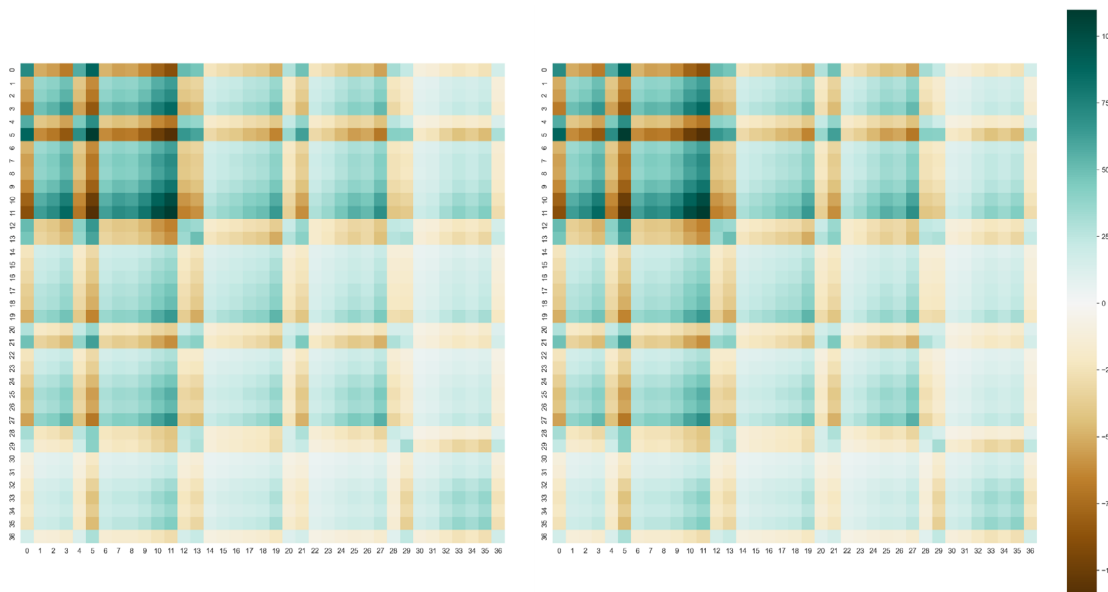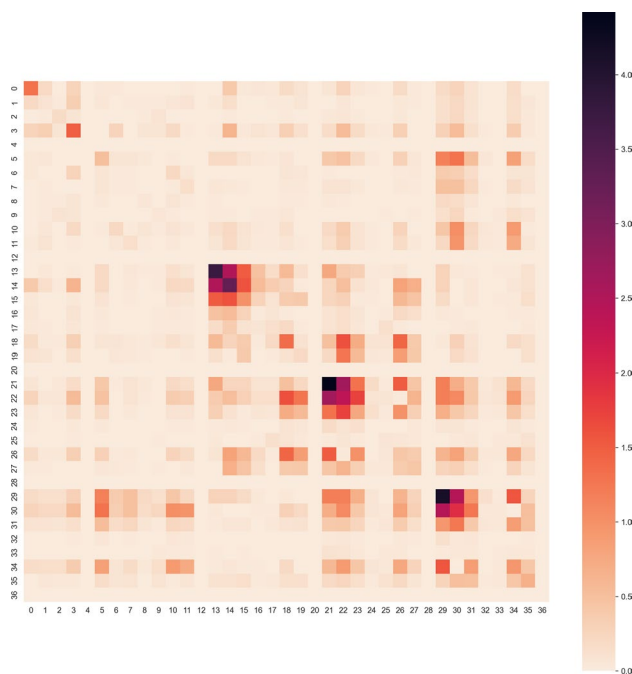

Fig. 1 A heatmap of the elements of the exact Hessian $\mathbf{H}$ (left) and Hessian approximant ${}^{PM7}\mathbf{H}$ (right), raised to the fifth root. The 37×37 Hessian contains second derivatives for the MNDO parameters $\alpha, \beta_s, \beta_p, U_{ss}, U_{pp}, \zeta_s, \zeta_p, E_{isol}$ for the elements H, C, N, O and F

Table 1 Eigenvalues obtained from diagonalization of the approximant $^{PM7}\mathbf{H}$ (left) and the exact Hessian $\mathbf{H}$ (right) using the parameters from the MNDO model. The exact reference data and weighting used to construct the error function is detailed in Table 2

No	Eigenvalue ($^{PM7}\mathbf{H}$)	Eigenvalue ($\mathbf{H}$)
1	0.128368	−512,816
2	0.934416928	−460,096.7651
3	15.41202367	−267,302.7955
4	30.05635772	−216,687.9532
5	61.02869235	−167,602.8167
6	89.83798894	−39,421.75011
7	133.3580448	−7503.560232
8	314.5683375	−2616.97078
9	333.6432758	−983.335745
10	1440.455065	−533.6458032
11	1645.025654	−211.742801
12	2550.646282	−101.2912023
13	3835.543251	−41.93718193
14	5680.70884	248.1694688
15	8639.457771	775.4405619
16	13,231.619	2083.752434
17	20,549.13091	3068.359634
18	38,971.27902	11,361.85088
19	103,579.4856	19,381.78711
20	121,144.8051	68,127.12467
21	182,104.4115	119,463.3803
22	247,015.4439	298,112.1443
23	313,366.6596	894,952.2348
24	536,170.167	1,204,079.117
25	755,156.5518	1,925,370.492
26	968,678.9509	2,107,406.892
27	1,153,754.339	2,604,117.089
28	2,096,012.221	8,556,379.32
29	3,620,532.328	8,932,577.125
30	5,794,531.367	17,840,724.68
31	12,084,613.12	38,505,843.45
32	38,639,528.07	72,107,717.07
33	187,190,499.4	190,046,809.8
34	446,920,022	450,868,384.7
35	509,186,668.9	534,566,784.3
36	854,723,315.4	862,489,769.1
37	47,330,423,728	47,359,707,042

Fig. 2 A heatmap of percentage errors of each Hessian element, calculated as $\left| \frac{^{PM7}\mathbf{H}_{ij} - \mathbf{H}_{ij}}{\mathbf{H}_{ij}} \right| \times 100\%$

$$\frac{d}{d^{Z_C}p}\left(\frac{\partial E_{AB}}{\partial^A \tau}\right) = \sum_{\mu,\nu \in A} \left(\frac{dP_{\mu\nu}}{d^{Z_C}p}\frac{dV_{\mu\nu,B}}{d^A\tau} + P_{\mu\nu}\frac{d^2 V_{\mu\nu,B}}{d^A\tau d^{Z_C}p}\right)$$
$$+ \sum_{\lambda,\sigma \in B}\left(\frac{dP_{\lambda\sigma}}{d^{Z_C}p}\frac{dV_{\lambda\sigma,A}}{d^A\tau} + P_{\lambda\sigma}\frac{d^2 V_{\lambda\sigma,A}}{d^A\tau d^{Z_C}p}\right) + 2\sum_{\substack{\mu \in A \\ \lambda \in B}}\left(\frac{dP_{\mu\lambda}}{d^{Z_C}p}\frac{d\beta_{\mu\lambda}}{d^A\tau} + P_{\mu\lambda}\frac{d^2\beta_{\mu\lambda}}{d^A\tau d^{Z_C}p}\right)$$
$$+ \sum_{\substack{\mu,\nu \in A \\ \lambda,\sigma \in B}}\left(P_{\mu\nu}P_{\lambda\sigma} - {}^\alpha P_{\mu\nu}{}^\alpha P_{\lambda\sigma} - {}^\beta P_{\mu\nu}{}^\beta P_{\lambda\sigma}\right)\frac{d^2(\mu\nu|\lambda\sigma)}{d^A\tau d^{Z_C}p}$$
$$+ \sum_{\substack{\mu,\nu \in A \\ \lambda,\sigma \in B}}\left(\frac{dP_{\mu\nu}}{d^{Z_C}p}P_{\lambda\sigma} + P_{\mu\nu}\frac{dP_{\lambda\sigma}}{d^{Z_C}p}\right)\frac{d(\mu\nu|\lambda\sigma)}{d^A\tau}$$
$$- \sum_{\substack{\mu,\nu \in A \\ \lambda,\sigma \in B}}\frac{d(\mu\nu|\lambda\sigma)}{d^A\tau}\sum_{\sigma \in \{\alpha,\beta\}}\left(\frac{d^\sigma P_{\mu\nu}}{d^{Z_C}p}{}^\alpha P_{\lambda\sigma} + {}^\alpha P_{\mu\nu}\frac{d^\sigma P_{\lambda\sigma}}{d^{Z_C}p}\right) \tag{56}$$

$$\frac{d}{d^{Z_C}p}\left(\sum_{B\neq A}\frac{dV_{AB}^{CRF}}{d^A\tau}\right) = \sum_{B\neq A}\frac{d^2 V_{AB}^{CRF}}{d^A\tau d^{Z_C}p} \tag{57}$$

Elements of $\frac{d^2\mathbf{g}}{d^{Z_C}p_i d^{Z_D}p_j}$ are likewise obtained by direct differentiation of the above expression and are omitted for brevity.

Further details on the equations in this section, where necessary, are provided in the Supplementary Information.

Nature of the Hessian approximant in PM7

To illustrate the differences between the exact Hessian $\mathbf{H}$ and the approximant $^{PM7}\mathbf{H}$, a pictorial representation

Lastly, the elements of $\frac{d\mathbf{g}}{d^{Z_C}p}$ (derivative of the gradient vector in Cartesian coordinates against arbitrary parameter ^{Z_C}p) are given by

$$\frac{d^2 E}{d^A\tau d^{Z_C}p} = \sum_{B\neq A}\frac{d}{d^{Z_C}p}\left(\frac{\partial E_{AB}}{\partial^A\tau}\right) + \frac{d}{d^{Z_C}p}\left(\sum_{B\neq A}\frac{dV_{AB}^{CRF}}{d^A\tau}\right) \tag{55}$$

of the two matrices is given in Fig. 1. The matrices were computed using original MNDO parameters [1] for a chosen training set of 1206 molecules (see Supplementary Information).

Furthermore, the approximant $^{PM7}\mathbf{H}$ is observed to be positive definite while $\mathbf{H}$ reflects that the parameter surface is non-convex (Table 1).

Methods for parameter optimization

In PM7, parameter optimization is performed via an approximated line search, with the search direction obtained via a direct Hessian descent (HD) step on the approximated Hessian [5]:

$$\hat{\mathbf{d}} = -\frac{^{PM7}\mathbf{H}^{-1}\mathbf{g}}{\left|^{PM7}\mathbf{H}^{-1}\mathbf{g}\right|}, \quad |\mathbf{d}| = \text{argmax}_k\left(\widetilde{\mathcal{S}}\big|_{\mathbf{p}=\mathbf{p}_0+k\hat{\mathbf{d}}}\right) \tag{58}$$

$$\widetilde{\mathcal{S}} = \sum_\alpha \mathcal{C}_\alpha^2 \left(\xi_\alpha^{\text{ref}} - \xi_\alpha\big|_{\mathbf{p}=\mathbf{p}_0} - \sum_{Z_A}\sum_{Z_A p_i}\frac{d\xi_\alpha}{d^{Z_A}p_i}\bigg|_{\mathbf{p}=\mathbf{p}_0}\left(^{Z_A}p_i\big|_{\mathbf{p}} - ^{Z_A}p_i\big|_{\mathbf{p}=\mathbf{p}_0}\right)\right)^2 \tag{59}$$

Table 2 Weighting factors $\mathcal{C}_i$ for the reference properties used in parameterization. The weighting factors for ΔH_f, *I.E.* and $\langle\mu\rangle$ are the same as in the PMx methods

Property	Units	$\mathcal{C}_i$
ΔH_f	kcal/mol	1mol/kcal
I.E.	eV	10eV^{-1}
$\langle\mu\rangle$	D	20D^{-1}
$|\mathbf{g}|$	kcal/(mol · bohr)	$0.5\text{mol} \cdot \text{bohr/kcal}$

This method of determining the step size shall be termed the approximated line search (ALS) method; an alternative would be a trust radius (TR), where the step size is constrained by a dynamic trust radius. The modification of the trust radius in our attempt is given by the following computation, with $\mathbf{B}$ representing an arbitrary Hessian or Hessian approximant:

$$Q_n = \mathbf{d}_n^T\mathbf{g}_n + \frac{1}{2}\mathbf{d}_n^T\mathbf{B}_n\mathbf{d}_n, \rho_n = \frac{\mathcal{S}|_{\mathbf{p}=\mathbf{p}_n+\mathbf{d}_n} - \mathcal{S}|_{\mathbf{p}=\mathbf{p}_n}}{Q_n} \tag{60}$$

$$|\mathbf{d}_n| = R_n, R_{n+1} = \begin{cases} \frac{5}{4}R_n, \rho_n > 0.8 \, and \, S|_{\mathbf{p}=\mathbf{p}_n+\mathbf{d}_n} < S|_{\mathbf{p}=\mathbf{p}_n} \\ \frac{1}{2}R_n, \rho_n < 0.25 \, or \, S|_{\mathbf{p}=\mathbf{p}_n+\mathbf{d}_n} \geq S|_{\mathbf{p}=\mathbf{p}_n} \\ R_n, otherwise \end{cases} \tag{61}$$

In addition to determination of $\hat{\mathbf{d}}$ via direct Hessian descent, a trust region optimization (TRO) was also attempted, where $\mathbf{d}$ is computed via:

$$\mathbf{d} = (\mathbf{B} - \lambda\mathbf{I})^{-1}\mathbf{g}, |\mathbf{d}| = R \tag{62}$$

The shift parameter λ is computed iteratively, with the chosen numerical method detailed in the Supplementary Information.

Lastly, three different choices for the second-derivative matrix $\mathbf{B}$ were investigated. In addition to the exact Hessian $\mathbf{H}$ and the approximant $^{PM7}\mathbf{H}$, a modified Hessian $\mathbb{H}$ was found to yield promising results:

$$\mathbb{H} = \mathbf{PD}'\mathbf{P}^{-1}, \mathbf{D}'_{ii} = \left|\mathbf{D}_{ii}\right|, \mathbf{H} = \mathbf{PDP}^{-1} \tag{63}$$

The modified Hessian $\mathbb{H}$ preserves the eigenvectors of the exact Hessian while converting it to be positive (semi)

Fig. 3 Optimization curve obtained when using the parameterization procedure detailed in [5] and [10]
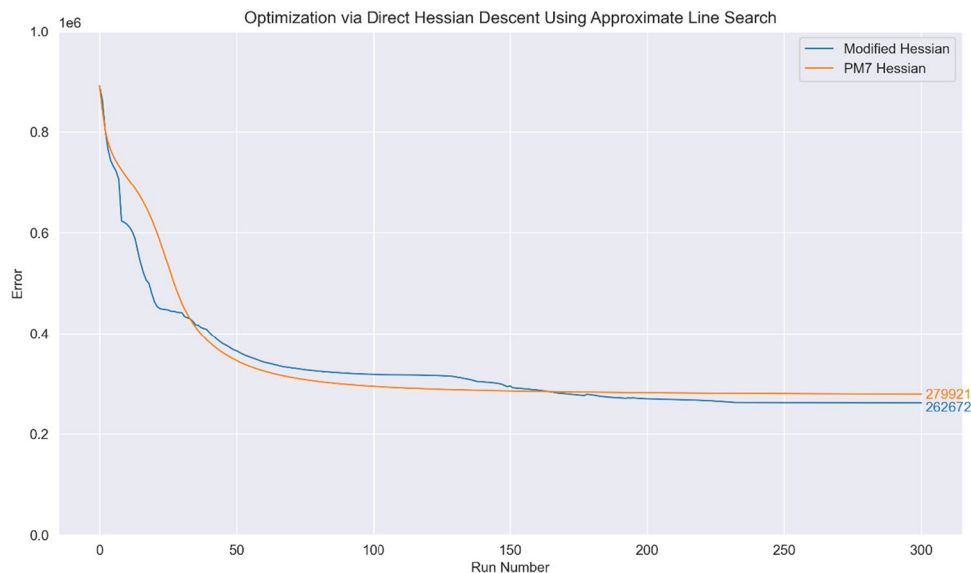
Fig. 4 Optimization curve obtained when using a trust radius with the direct Hessian descent (Newton–Raphson) step. The initial step size at run no. 0 is set at 2.0 for $\mathbb{H}$ and $^{PM7}\mathbf{H}$
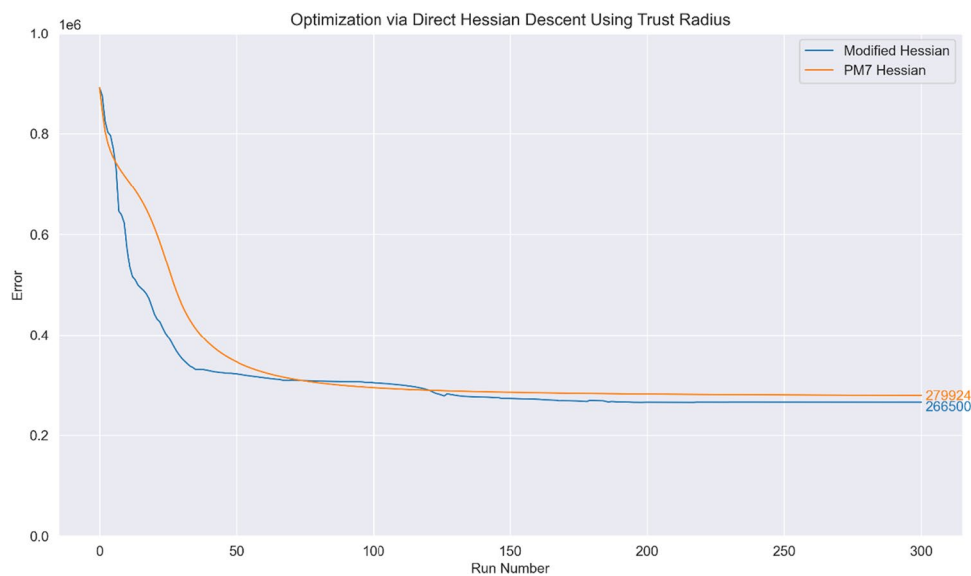

definite, and ensures that a direct Hessian descent step will not traverse in the uphill direction given a surface of the wrong concavity.

It should be noted that $^{PM7}\mathbf{H}$ is constructed without any second derivatives of molecular properties, and hence, optimization with $^{PM7}\mathbf{H}$ is comparable to other methods using approximate Hessians that do not evaluate second derivative information (e.g. the BFGS or DFP schemes); however, $^{PM7}\mathbf{H}$ appears to provide a reasonably good estimate for $\mathbf{H}$ despite having incorrect eigenvalue information (as seen in Fig. 2, the percentage errors for specific elements in the Hessian matrix are at most 4–5%) and is thus expected to perform better than a regular quasi-Newton

optimizer. We note in passing that PM3 and PM6 were optimized with the method outlined in [5] using a DFP update scheme while the more recent PM7 [10] was optimized using $^{PM7}\mathbf{H}$, suggesting that $^{PM7}\mathbf{H}$ better approximates the exact Hessian $\mathbf{H}$.

The computation of the gradient $\mathbf{g}$ is straightforward and may be obtained easily by both finite difference and analytic differentiation; this suggests that gradient-free optimization methods would not be necessary in optimizing parameters for NDDO-based methods. Such local optimizers can, however, be combined with other methods to identify global minima, e.g. genetic algorithms [30, 31], and may be indicative of possible avenues for further work on this subject. We

Fig. 5 Optimization curve obtained when using a trust region optimizer. The initial step size at run no. 0 is set at 0.1 for both curves
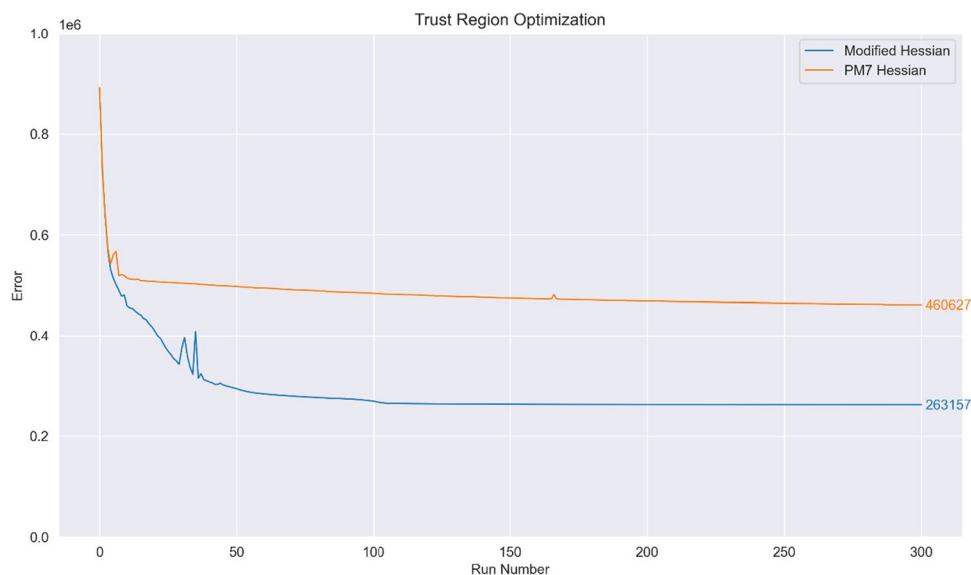
Table 3 Eigenvalues taken from diagonalizing the exact Hessian using parameters from run no. 300 of optimization with $^{PM7}\mathbf{H}$ (corresponding to the end point of the green curve in Fig. 5)

No	Eigenvalue	No	Eigenvalue	No	Eigenvalue
1	-41.234	14	6818.55467	27	1,512,211.907
2	-3.823761	15	8866.28568	28	2,277,598.426
3	-2.9794242	16	19,648.5226	29	4,749,192.867
4	7.18597596	17	34,039.3551	30	11,509,664.69
5	42.2832925	18	42,849.9713	31	20,200,473.71
6	53.6581437	19	122,256.195	32	37,260,138.2
7	114.412963	20	166,404.131	33	146,715,815.1
8	351.125533	21	219,075.729	34	359,303,765.9
9	778.324332	22	233,887.679	35	404,915,300.3
10	923.225124	23	328,171.073	36	771,116,378.5
11	1804.53062	24	499,506.976	37	39,386,113,276
12	3531.45311	25	878,679.923		
13	4324.88892	26	1,287,823.88		

defer a discussion of the nature of local and global minima on the parameter surface to the "Conclusion".

Results of limited parameterisation

As a proof of concept, a limited parameterization of 1206 molecules consisting of the atom types C, H, N, O, F was performed using the MNDO formalism; the relevant routines for the semi-empirical evaluation of molecular properties using MNDO were correspondingly implemented and compared against MOPAC [32] for accuracy (see the Supplementary Information). Geometrical data, e.g., bond angles or bond lengths, are accounted for in our parameterization procedure by using the norm of the gradient vector $|\mathbf{g}|$ at

a reference geometry (either experimental or the result of high-level calculations), with the corresponding reference function set to zero to simulate a perfect correspondence between the semi-empirical and reference geometries; this was chosen to facilitate the ease of preparing the relevant inputs. The weighting functions $\mathcal{C}_i$ for the error function were chosen as specified in Table 2.

While additional properties such as reaction barriers were also viable choices for inclusion as reference data in our training set, such information was not used in our limited parameterization; nonetheless, the parameter derivatives of the molecular energy used in calculating derivatives of ΔH_f can be applied to calculate the relevant parameter derivatives for the energies of transition states, allowing for an easy extension if necessary.

For all optimization methods, the graphs shown below are terminated once the decrease in $\mathcal{S}$ is no longer appreciable; the optimization runs discussed in this section hence do not represent the identification of minima on the parameter surface.

First, application of the optimization method used in the PMx models (Hessian descent with step size determined by the approximate line search) leads to surprisingly similar behaviour when both $\mathbb{H}$ and $^{PM7}\mathbf{H}$ are used; nonetheless, there is an appreciable difference in the final parameter values for the two methods. Direct employment of the Hessian matrix $\mathbf{H}$ results in a poor optimisation procedure as the evaluated step size shrinks significantly around $\mathcal{S} = 575000$; this is expected as the line-search procedure seeks to minimise $\mathcal{S}$ even as the Newton–Raphson step traverses uphill for nonconvex regions. The optimization curves for $\mathbf{H}$ are hence omitted in Figs. 3, 4 and 5.

To compare the line-search method used in the PMx models with a trust radius method, optimization using direct Hessian descent but with the step size determined via a

Fig. 6 Optimization curves obtained when using a trust region optimizer and the modified Hessian $\mathbb{H}$. The initial step size at run no. 0 is set at 0.1 for all three curves, but the maximum step size $|\mathbf{d}|_{max}$ was varied
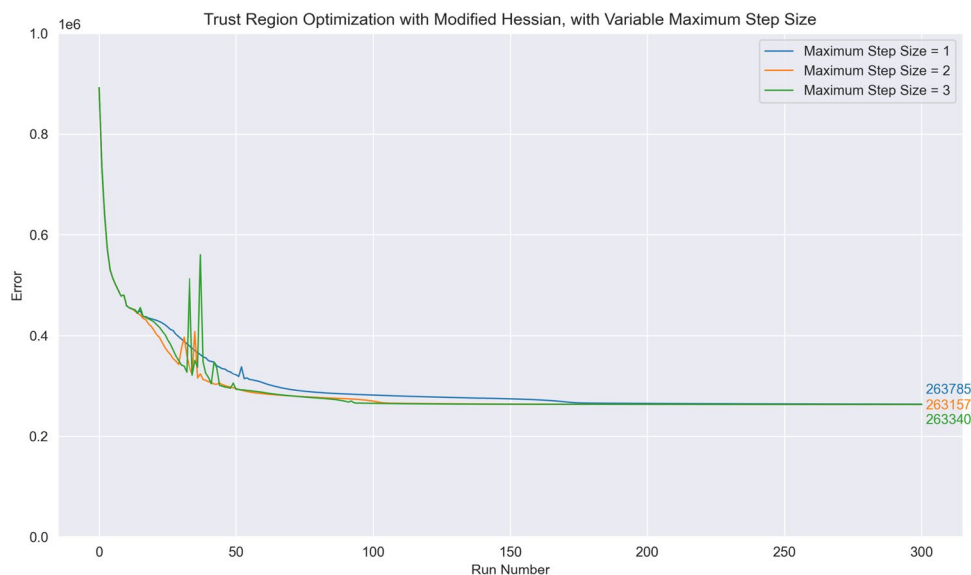
Table 4 Top and bottom: parameter values and their associated elements of the gradient vector used in parameter optimization for the parameters obtained at the end of our limited optimization

Parameter	Element				
	C	H	N	O	F
$z_A \alpha$	2.660499	2.993408	3.021163	3.334554	3.590480
$z_A \beta_s$	− 13.719865	− 7.334656	− 17.522551	− 73.616234	− 1753.340685
$z_A \beta_p$	− 6.970837	NIL	− 15.550446	− 22.327739	− 22.382738
$z_A U_{ss}$	− 47.292652	− 10.314472	− 64.905273	− 94.106196	− 107.380890
$z_A U_{pp}$	− 40.221274	NIL	− 57.638287	− 78.565181	− 106.952003
$z_A \zeta_s$	2.326858	1.124161	2.914488	5.594470	50.167806
$z_A \zeta_p$	1.637021	NIL	2.126141	2.425014	2.463961
$z_A E_{isol}$	− 112.533538	− 11.423363	− 189.262258	− 309.024988	− 434.411284
Gradient	Element				
	C	H	N	O	F
$z_A \alpha$	− 5631.2989	− 122.511277	− 265.68094	− 170.18825	− 1319.4913
$z_A \beta_s$	122.967088	39.777392	4.0270245	0.92212293	1.65447933
$z_A \beta_p$	284.128795	NIL	14.8445277	14.6824465	110.38728
$z_A U_{ss}$	264.36961	51.0183803	29.0786636	40.8640734	655.636142
$z_A U_{pp}$	655.367166	NIL	55.1677978	94.2737931	2154.95912
$z_A \zeta_s$	1724.84129	97.0883125	57.8930721	36.7178843	86.2475113
$z_A \zeta_p$	4822.82283	NIL	297.807752	334.865333	1932.25505
$z_A E_{isol}$	− 245.1164	− 66.13715	− 17.695251	− 21.930875	− 391.28695

Table 5 Eigenvalues taken from diagonalizing the exact Hessian using the parameters obtained at the end of our limited parameterization

No	Eigenvalue	No	Eigenvalue	No	Eigenvalue
1	0.000307191	14	7097.672134	27	1,169,583.719
2	2.110409103	15	11,248.65204	28	4,374,828.93
3	9.894574497	16	13,983.01	29	4,455,736.981
4	39.09862218	17	27,423.51093	30	9,638,553.272
5	88.84495043	18	31,607.2681	31	21,918,237.20
6	120.8695565	19	87,137.37687	32	32,920,532.04
7	418.3575145	20	180,143.5721	33	104,814,193.86
8	460.8171646	21	209,812.6169	34	241,208,268.54
9	755.3529097	22	324,571.6458	35	294,810,061.12
10	1256.25836	23	416,778.2783	36	555,086,343.83
11	2304.566797	24	555,533.5715	37	33,782,858,461.66
12	3297.414171	25	688,452.7501		
13	3864.018903	26	1,078,765.379		

dynamic trust radius was attempted. We note in passing that the resultant optimization curve strongly resembles that of the according line-search method (cf. Figure 3); once again, direct employment of **H** leads to very poor optimisation, and the resultant curve has been omitted:

Lastly, optimization was attempted with a trust region optimizer; this led to similar performance for $\mathbb{H}$ while ^{PM7}**H** performed markedly poorer. Optimization with **H** led to suboptimal parameters that encountered problems with SCF convergence and geometry optimization and was hence abandoned.

Optimization with ^{PM7}**H** appears to result in the identification of a saddle point, which is incorrectly identified as a minimum due to the positive-definite nature of ^{PM7}**H**; the Hessian eigenvalues (from the exact Hessian **H**) for the resultant parameters obtained via optimization with ^{PM7}**H** are given in Table 3.

In all methods (Hessian descent with both line search and trust radius determination of step size as well as the trust-region optimizer), the step size is constrained to a maximum of $|\mathbf{d}|_{max} = 2$ to ensure that there are no significant and unexpected increases in the error function. We note in passing that the choice of $|\mathbf{d}|_{max} = 2$, while arbitrary, does not play a significant role in the optimization of parameters; choices of $|\mathbf{d}|_{max} = 1$ and $|\mathbf{d}|_{max} = 3$ lead to very similar optimization curves (Fig. 6).

In the course of parameterization, it was realized that identification of a true local minimum would be difficult and require manual intervention in the parameterization procedure; thus, any set of parameters with a reasonably small gradient vector magnitude, a positive definite (exact) Hessian matrix and a marked resistance to further reduction in the error function should be accepted as a reasonable result from parameter optimization.

Fig. 7 Optimization curves obtained when using a trust region optimizer and the modified Hessian $\mathbb{H}$. The initial step size at run no. 0 is set at 0.1 for all three curves and $|\mathbf{d}|_{max} = 2$, but the initial parameters used were varied (see Tables 6 and 7)
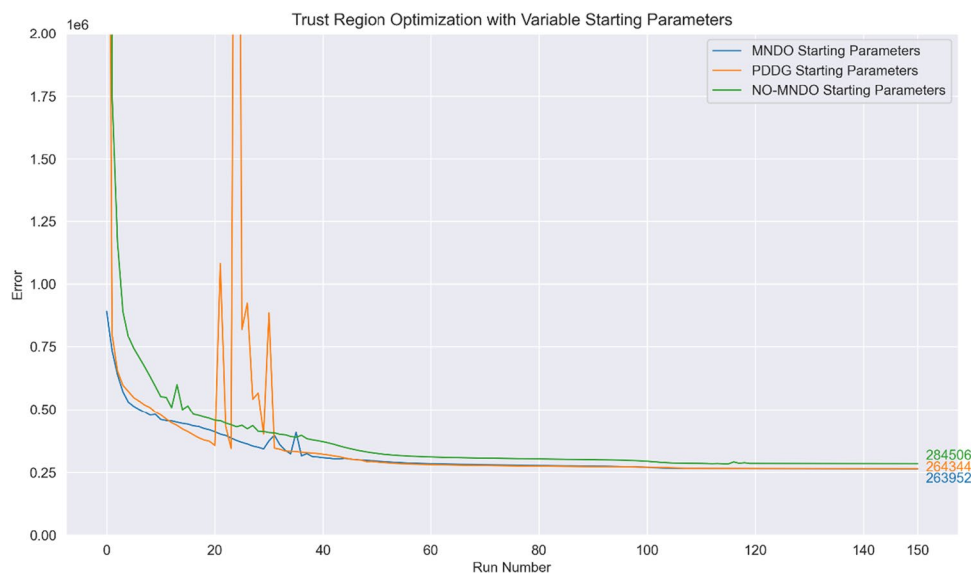
Table 6 Initial parameter values obtained from PDDG/MNDO (see Fig. 7)

Parameter	Element				
	C	H	N	O	F
$^{Z_A}\alpha$	2.555522	2.491813	2.843678	3.238842	3.322382
$^{Z_A}\beta_s$	−18.841334	−7.493504	−20.375774	−33.606336	−67.827612
$^{Z_A}\beta_p$	−7.922234	NIL	−21.085373	−27.984442	−40.924818
$^{Z_A}U_{ss}$	−53.837582	−11.724114	−71.871894	−97.88497	−134.22038
$^{Z_A}U_{pp}$	−39.936409	NIL	−58.216617	−77.342674	−107.15596
$^{Z_A}\zeta_s$	1.809817	1.322431	2.231424	2.569172	4.328519
$^{Z_A}\zeta_p$	1.825008	NIL	2.25346	2.697152	2.905042
$^{Z_A}E_{isol}$	−123.86441	−12.015956	−206.46663	−310.87975	−488.70324

Table 7 Initial parameter values obtained from NO-MNDO (see Fig. 7)

Parameter	Element				
	C	H	N	O	F
$^{Z_A}\alpha$	2.48446	2.687705	2.658599	2.946645	3.4196606
$^{Z_A}\beta_s$	−16.208034	−9.364858	−24.90552	−35.477596	−48.290466
$^{Z_A}\beta_p$	−10.637421	NIL	−21.291958	−28.881783	−36.50854
$^{Z_A}U_{ss}$	−50.189763	−10.880363	−69.782951	−96.705658	−131.07155
$^{Z_A}U_{pp}$	−39.547267	NIL	−56.981889	−76.391762	−105.78214
$^{Z_A}\zeta_s$	1.925428	1.061597	2.351138	2.455548	2.848487
$^{Z_A}\zeta_p$	1.727933	NIL	1.951819	2.537964	2.848487
$^{Z_A}E_{isol}$	−119.5944	−13.160122	−202.2436	−304.34129	−476.68378

In this work, we report the identification of a local minimum from our limited parameterization, obtained with the imposition of no constraints on the parameters. The parameters, as well as the values of the associated gradient element, for an identified minimum with $\mathcal{S} = 261996$ are given in Table 4; the associated Hessian eigenvalues are presented in Table 5.

While clearly unphysical, these parameters indicate a position close to a local minimum on the parameter surface

Table 8 Top, middle and bottom: parameter values after 150 optimization runs starting from MNDO parameters, PDDG/MNDO parameters and NO-MNDO parameters (see Tables 6 and 7)

MNDO parameter	Element				
	C	H	N	O	F
$z_A \alpha$	2.65893028	2.98616999	3.01789735	3.33094136	3.59661896
$z_A \beta_s$	-13.896133	-7.3703509	-17.576889	-72.876223	-219.98053
$z_A \beta_p$	-6.9485023	NIL	-15.611441	-22.291261	-20.757317
$z_A U_{ss}$	-47.280029	-10.315317	-64.95198	-94.806035	-107.24866
$z_A U_{pp}$	-40.191348	NIL	-57.601922	-78.551712	-106.91534
$z_A \zeta_s$	2.33907785	1.12519887	2.90566782	5.48159077	12.7783341
$z_A \zeta_p$	1.62963659	NIL	2.12324781	2.42135448	2.39933061
$z_A E_{isol}$	-112.50774	-11.42606	-189.30868	-310.3944	-434.01522
PDDG/MNDO parameter	Element				
	C	H	N	O	F
$z_A \alpha$	2.66899203	2.92239613	3.02052461	3.33649311	3.58643208
$z_A \beta_s$	-13.972814	-7.364351	-16.728601	-72.368252	-230.56023
$z_A \beta_p$	-6.8267709	NIL	-15.800279	-22.29721	-21.980605
$z_A U_{ss}$	-47.306366	-10.336626	-64.273955	-95.398437	-107.30999
$z_A U_{pp}$	-40.14483	NIL	-57.523819	-78.507984	-106.77862
$z_A \zeta_s$	2.35713045	1.18520195	2.94159025	5.44939528	13.2748539
$z_A \zeta_p$	1.62332729	NIL	2.12477097	2.42716075	2.46038129
$z_A E_{isol}$	-112.36726	-11.193921	-187.80396	-311.39522	-433.37951
NO-MNDO parameter	Element				
	C	H	N	O	F
$z_A \alpha$	2.66382164	2.98279164	3.03799145	3.35980247	3.71454355
$z_A \beta_s$	-13.386239	-7.3096246	-18.540266	-82.349631	-185.91367
$z_A \beta_p$	-6.999786	NIL	-15.011623	-21.616282	-16.258881
$z_A U_{ss}$	-47.243051	-10.247595	-65.176426	-92.97953	-105.84392
$z_A U_{pp}$	-40.187542	NIL	-57.741407	-78.760165	-107.38397
$z_A \zeta_s$	2.30677182	1.14028611	3.00201012	6.13082666	12.0426399
$z_A \zeta_p$	1.64820504	NIL	2.09568357	2.38054219	2.14285707
$z_A E_{isol}$	-112.2377	-11.304432	-190.16538	-307.79559	-434.22304

and are an indication of the significant deficiencies of the MNDO model in modelling charged or radical species.

Notably, ammonia is predicted to be planar using these parameters; as the non-planarity of ammonia was not a constraint imposed upon the system for parameterization, this is not an erroneous result. A detailed tabulation of the predicted molecular properties at the given parameters is provided in the Supplementary Information.

In Fig. 7, optimization using $\mathbb{H}$ with initial step size 0.1 and $|\mathbf{d}|_{max} = 2$ is compared for three sets of starting parameters: the original MNDO parameters, the MNDO parameters reported for PDDG/MNDO [11] and the parameters for the elements C, H, N and O reported in NO-MNDO [22] alongside original MNDO parameters

for fluorine (see Tables 6 and 7). Since the formalism used remains that of MNDO and not PDDG/MNDO or NO-MNDO, these additional sets of parameters are expected to perform worse than the original MNDO parameters; however, they provide an interesting case study for how our local optimizer might behave when starting from different initial parameter values. The graph is terminated after 150 optimization runs, and the final error function values obtained are hence not reflective of the identification of a local minimum; nonetheless, the very similar error function values despite the significant differences in final parameters (as reported in Table 8) seem to indicate that there may be numerous local minima with similar error function values.

Conclusion

In this paper, we report a fully analytic differentiation routine for the evaluation of parameter derivatives for construction of the parameter gradient and Hessian in MNDO-based semi-empirical methods and have applied these equations for a proof-of-concept optimization of the MNDO parameters for the elements C, H, N, O and F. While the PM7 Hessian $^{PM7}\mathbf{H}$ appears to work remarkably well in parameterization schemes, it does not guarantee that the identified stationary point on the parameter surface will be a local minimum; furthermore, $^{PM7}\mathbf{H}$ does appear to perform worse when the optimization nears a local minimum or stationary point. The full Hessian $\mathbf{H}$ may be necessary for optimization near the stationary point, and the accurate eigenvalue information may also be helpful in ensuring that a local minimum is attained at the termination of parameterization.

We have placed significant emphasis on the identification of local minima on the parameter surface. While it is conceded that identification of global minima is of greater concern in parameterization, we note that parameter surfaces encountered in the optimization of neural networks often show multiple local minima, all with similar loss function values [33] and posit that a similar situation may be observed in NDDO-based methods.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00894-023-05499-3.

Author contribution A.O. wrote the main manuscript text and S.C. prepared all figures and tables in the manuscript. K.L.C. provided guidance and supervised the research work. All authors reviewed the manuscript.

Data Availability The relevant data in this paper shall be provided upon request.

Declarations

Ethical approval Not applicable.

Competing interests Not applicable.

References

1. Dewar MJS, Thiel W (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. J Am Chem Soc 99:4899–4907

2. Dewar MJ, Zoebisch EG, Healy EF, Stewart JJ (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. J Am Chem Soc 107(13):3902–3909

3. Rocha GB, Freire RO, Simas AM, Stewart JJ (2006) RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. J Comput Chem 27:1101–1111

4. Thiel W, Voityuk AA (1992) Extension of the MNDO formalism to d orbitals: integral approximations and preliminary numerical results. Theoret Chim Acta 81:391–404

5. Stewart JJP (1989) Optimization of parameters for semiempirical methods I Method. J Comput Chem 10:209–220

6. Stewart JJ (1989) Optimization of parameters for semiempirical methods II. Appl J Comput Chem 10:221–264

7. Stewart JJ (1991) Optimization of parameters for semiempirical methods III extension of PM3 to Be, Mg, Zn, Ga, Ge, as, SE, CD, in, Sn, Sb, Te, Hg, Tl, Pb, and Bi. J Comput Chem 12:320–341

8. Stewart JJ (2004) Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. J Mol Model 10:155–164

9. Stewart JJP (2007) Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. J Mol Model 13:1173–1213

10. Stewart J (2012) Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model 19(1):1–32

11. Repasky MP, Chandrasekhar J, Jorgensen WL (2002) PDDG/PM3 and PDDG/MNDO: improved semiempirical methods. J Comput Chem 23:1601–1622

12. Husch T, Vaucher AC, Reiher M (2018) Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. Int J Quantum Chem 118:e25799

13. Zhang H, Lau JWZ, Wan L, Shi L, Shi Y, Cai H, Luo X, Lo GQ, Lee CK, Kwek LC, Liu AQ, (2023) Molecular Property Prediction with Photonic Chip-Based Machine Learning. Laser Photonics Rev 17:2200698

14. Bhat HS, Ranka K, Isborn CM (2020) Machine learning a molecular Hamiltonian for predicting electron dynamics. Int J Dyn Control 8(4):1089–1101

15. Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA (2015) Big Data meets quantum chemistry approximations: the Δ-machine learning approach. J Chem Theory Comput 11(5):2087–2096

16. Chen J, Xu W, Zhang R (2021) A machine learning approach using frequency descriptor for molecular property predictions. New J Chem 45(44):20672–20680

17. Dral PO, von Lilienfeld OA, Thiel W (2015) Machine learning of parameters for accurate semiempirical quantum chemical calculations. J Chem Theory Comput 11(5):2120–2125

18. Kolb M, Thiel W (1993) Beyond the MNDO model: methodical considerations and numerical results. J Comput Chem 14:775–789

19. Weber W, Thiel W (2000) Orthogonalization corrections for semiempirical methods. Theor Chem Acc 103:495–506

20. Scholten M (2003) Semiemirische Verfahren mit Orthogonalisierungskorrek- turen: Die OM3 Methode. Heinrich-Heine-Universität Düsseldorf, Thesis

21. Dral PO, Wu X, Spörkel L, Koslowski A, Weber W, Steiger R, Scholten M, Thiel W (2016) Semiempirical quantum-chemical orthogonalization-corrected methods: theory, implementation, and parameters. J Chem Theory Comput 12:1082–1096

22. Sattelmeyer KW, Tubert-Brohman I, Jorgensen WL (2006) NO-MNDO: reintroduction of the overlap matrix into MNDO. J Chem Theory Comput 2:413–419

23. Goeppert-Mayer M, Sklar AL (1938) Calculations of the lower excited levels of benzene. J Chem Phys 6:645–652

24. Dewar MJS, Thiel W (1976) A semiempirical model for the two-center repulsion integrals in the NDDO approximation. Theor Chim Acta 46:89–104

25. Pulay P (1969) Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules. Molecular Physics 17(2):197–204

26. Patchokovskii S, Thiel W (1996) Analytical second derivatives of the energy in MNDO methods. J Comput Chem 17(11):1318–1327

27. Frisch M, Scalmani G, Vreven T, Zheng G (2009) Analytic second derivatives for semiempirical models based on MNDO. Mol Phys 107:881–887

28. Osamura Y, Yamaguchi Y, Schaefer HF (1986) Second-order coupled perturbed Hartree-Fock equations for closed-shell and open-shell self-consistent-field wavefunctions. Chem Phys 103:227–243

29. Pople JA, Krishnan R, Schelegel HB, Binkley JS (1979) Derivative studies in Hartree-Fock and Møller-Plesset theories. Intl J Quantum Chem 13:225

30. Rossi I, Truhlar DG (1995) Parameterization of NDDO wavefunctions using genetic algorithms an evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. Chem Phys Letters 233(3):231–236

31. Brothers EN, Merz KM (2002) Sodium parameters for AM1 and PM3 optimized using a modified genetic algorithm. J Phys Chem B 106(10):2779–2785

32. Stewart JJ (1990) Mopac: A semiempirical molecular orbital program. J Comput Aided Mol Des 4:1–103

33. Choromańska A, Henaff M, Mathieu M, Arous GB, LeCun Y (2014) The loss surfaces of multilayer networks. International Conference on Artificial Intelligence and Statistics