JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Trust Region Minimization of Orbital Localization Functions

Ida-Marie Høyvik,* Branislav Jansik, and Poul Jørgensen

qLEAP, Dept. of Chemistry, Aarhus University, Langelandsgade 140, DK-8000 Aarhus, Denmark

ABSTRACT: The trust region method has been applied to the minimization of localization functions, and it is shown that both local occupied and local virtual Hartree−Fock (HF) orbitals can be obtained. Because step sizes are size extensive in the trust region method, large steps may be required when the method is applied to large molecular systems. For an exponential parametrization of the localization function only small steps are allowed, and the standard trust radius update therefore has been replaced by a scheme where the direction of the step is determined using a conservative estimate of the trust radius and the length of the step is determined from a line search along the obtained direction. Numerical results for large molecular systems have shown that large steps can then safely be taken, and a robust and nearly monotonic convergence is obtained.

1. INTRODUCTION

During the past 50 years a vast number of orbital localization schemes have been presented. The key feature in these schemes is to use the property that the Hartree−Fock (HF) energy is invariant with respect to a unitary transformation among the occupied and among the virtual orbitals to determine the transformation that gives local HF orbitals. The most commonly used localization schemes are the ones of Boys,[1−3] Edmiston and Ruedenberg,[4,5] and Pipek and Mezey.[6] These schemes have been successfully used to obtain local occupied HF orbitals using a Jacobi sweep of iterations.[4] The localization of virtual orbitals is more difficult, and localizations of the virtual HF orbitals have in general failed for the commonly used localization schemes when a Jacobi sweep of iterations is used. To improve locality of the least local orbitals Jansik et al.[7] recently introduced a localization scheme where a penalty was imposed on the least local orbitals by minimizing powers of the orbital variance. Without a penalty this scheme is equivalent to the Boys localization. Jansik et al. further showed that both local occupied and local virtual HF orbitals can be obtained when a trust region method is used instead of a Jacobi sweep of iterations. We here describe in more detail an implementation of the trust region method for orbital localization.

Trust region minimization[8] is more involved than a Jacobi sweep of iterations as it requires local information about the gradient and Hessian. For that reason the method is also much more stable and robust. In the trust region method a trust region is defined where the objective function is approximated by a quadratic function. The size of the trust region is dynamically updated during the optimization according to the quadraticity of the function surface, and steps on the function surface are only trusted if the function surface is close to quadratic in the step region

Step norms are size extensive in the trust region method. This means that when the molecular size increases the step norms increase. For an exponential parametrization of the localization function the second order Taylor expansion is uniquely defined only for one period of the exponential function. Further, for the exponential function to be dominated by its linear term, the individual components of the exponential function have to be small (less than $\pi/4$).[9] For large molecular systems large step norms may be encountered. These may be due either to a sum of many small components or to one single large element. In the first case the trust region method can straightforwardly be applied, whereas in the second case the step has to be severely reduced for the step to be trusted. Since the structure of a step is not known a priori and since we want to be able to take large steps when the trust region method is applied, we have developed an alternative scheme to the trust region update. In this scheme the step direction is determined from a conservative estimate of the trust region without update of the trust region, and a line search is afterward performed along the obtained direction to determine the optimal length of the step. In this way very large downhill steps may safely be taken.

The trust region minimization method was developed by Fletcher and introduced into quantum chemistry by Swanstrøm et al.[10] for the optimization of multiconfigurational self-consistent field (MCSCF) wave functions. In the form of solving an augmented Hessian eigenvalue problem it was used by Jensen et al.,[9] for optimizing MCSCF wave functions and by Salek et al.[11] for optimizing SCF and Kohn−Sham (KS) densities. The augmented Hessian approach with no step length control was introduced by Lengsfield for MCSCF wave functions.[12] Cerjan and Miller[13] analyzed the second order expansion for the determination of stationary points on potential energy surfaces without step restrictions, and Simons and Jørgensen[14] further developed this method into a restricted step method.

In Section 2 the trust region minimization method is described in a context where the level shifted Newton equation is solved as an eigenvalue problem. In Section 3 convergence results are presented, followed by a summary and concluding remarks in Section 4.

2. TRUST REGION MINIMIZATION

We will now describe the most important aspects of the trust region optimization as implemented for the optimizations of localization functions and refer to Fletcher[8] for a more

Received: June 8, 2012
Published: August 17, 2012

   dx.doi.org/10.1021/ct300473g | J. Chem. Theory Comput. 2012, 8, 3137−3146

elaborate and general discussion on the trust region method. The trust region implementation we describe follows an outline similar to the one which was used for optimizing MCSCF wave functions.[9] However, in our implementation the trust region update is replaced with an algorithm where the direction of the step is obtained from a conservative estimate of the trust region and the length of the step from a line search along this direction.

2.1. Trust Region Method. To obtain a set of local Hartree−Fock orbitals we have to minimize a localization function ξ. The orbitals $\{|p\rangle,|q\rangle\cdots\}$ (e.g., a set of occupied or a set of virtual HF orbitals) may be parametrized in terms of an orthogonal transformation between the orbitals

$$|p\rangle = \sum_q |q\rangle[\exp(-\boldsymbol{\kappa})]_{qp} \tag{1}$$

where $\boldsymbol{\kappa}$ is an antisymmetric matrix of parameters defined as[15]

$$\boldsymbol{\kappa} = \sum_{p>q} \kappa_{pq}(\boldsymbol{E}_{pq} - \boldsymbol{E}_{qp}), \quad [\boldsymbol{E}_{pq}]_{rs} = \delta_{pr}\delta_{qs} \tag{2}$$

The trust region optimization methods takes its starting point in a second order Taylor series

$$\Lambda(\boldsymbol{\kappa}) = \xi^{[0]} + \boldsymbol{\kappa}^T\boldsymbol{\xi}^{[1]} + \frac{1}{2}\boldsymbol{\kappa}^T\boldsymbol{\xi}^{[2]}\boldsymbol{\kappa} \tag{3}$$

where $\xi^{[0]}$ is the value of ξ at the expansion point and $\boldsymbol{\xi}^{[1]}$ and $\boldsymbol{\xi}^{[2]}$ are the gradient and Hessian of ξ respectively with $\boldsymbol{\kappa} = \mathbf{0}$ as expansion point. The basic idea in trust region optimization is to take steps only in the region where the second order expansion is a good approximation to the full expansion. This region is termed the trust region and is a hypersphere of radius h around the expansion point.

For a minimization, a Newton step

$$\boldsymbol{\xi}^{[2]}\boldsymbol{\kappa} = -\boldsymbol{\xi}^{[1]} \tag{4}$$

is taken if the Hessian is positive definite and the step is inside the trust region. This step determines the minimum of $\Lambda(\boldsymbol{\kappa})$ and a sequence of iterations will converge quadratically to the minimum. In contrast, if the step goes beyond the trust region, or if the Hessian is not positive definite, the step giving the minimum value of $\Lambda(\boldsymbol{\kappa})$ on the boundary of the trust region has to be determined. This step is determined by setting up the Lagrangian

$$L(\boldsymbol{\kappa}, \mu) = \xi^{[0]} + \boldsymbol{\kappa}^T\boldsymbol{\xi}^{[1]} + \frac{1}{2}\boldsymbol{\kappa}^T\boldsymbol{\xi}^{[2]}\boldsymbol{\kappa} - \frac{1}{2}\mu(\boldsymbol{\kappa}^T\boldsymbol{\kappa} - h^2) \tag{5}$$

where μ is the undetermined multiplier for the constraint

$$\boldsymbol{\kappa}^T\boldsymbol{\kappa} = h^2 \tag{6}$$

The stationary points of the Lagrangian are determined by solving the level shifted Newton-equations.

$$(\boldsymbol{\xi}^{[2]} - \mu\mathbf{1})\boldsymbol{\kappa}(\mu) = -\boldsymbol{\xi}^{[1]} \tag{7}$$

Many stationary points exists on the boundary of the hypersphere of radius h. In the trust region method the step which determine the minimum is chosen. The minimizing step may uniquely be determined by imposing constraints on the allowed values of multiplier μ in eq 5. To understand how this may be performed we first recognize that the step norm in eq 6 is invariant with respect to a coordinate transformation. We may therefore examine the step length function in the diagonal representation of $\boldsymbol{\xi}^{[2]}$

$$\boldsymbol{\varepsilon} = \mathbf{U}^\dagger\boldsymbol{\xi}^{[2]}\mathbf{U} \tag{8}$$

$$\boldsymbol{X}(\mu) = \boldsymbol{U}^\dagger\boldsymbol{\kappa}(\mu), \quad \boldsymbol{g} = \boldsymbol{U}^\dagger\boldsymbol{\xi}^{[1]} \tag{9}$$

where the step length function becomes

$$\|\boldsymbol{X}(\mu)\|^2 = -\sum_i (\varepsilon_i - \mu)^{-2} g_i^2 \tag{10}$$

To identify the appropriate values of μ consider the first and second order change in ξ.

$$\Delta\xi_1 = \boldsymbol{\xi}^{[1]^T}\boldsymbol{\kappa} = \sum_i (\mu - \varepsilon_i)\left(\frac{g_i}{\varepsilon_i - \mu}\right)^2 \tag{11}$$

$$\Delta\xi_2 = \boldsymbol{\xi}^{[1]^T}\boldsymbol{\kappa} + \frac{1}{2}\boldsymbol{\kappa}^T\boldsymbol{\xi}^{[2]}\boldsymbol{\kappa} = \sum_i \left(\mu - \frac{1}{2}\varepsilon_i\right)\left(\frac{g_i}{\varepsilon_i - \mu}\right)^2 \tag{12}$$

For a minimization all individual contributions from all eigenmodes in eq 11 and eq 12 must be negative, which is obtained selecting $\mu < \varepsilon_1$.

In Figure 1 (top) we give an example of a step length function $\|X(\mu)\| = \|\boldsymbol{\kappa}(\mu)\|$ for a 100-dimensional model system plotted in the region $-8 < \mu < 1$. The step length function is positive with asymptotes at the Hessian eigenvalues which are marked with vertical dotted lines in Figure 1. The lowest Hessian eigenvalue is −5.2 in Figure 1 (top). Since $\|X(\mu)\|$ is monotonically increasing from 0 to ∞ for $-\infty < \mu < \varepsilon_1$ the minimizing step is uniquely determined. The

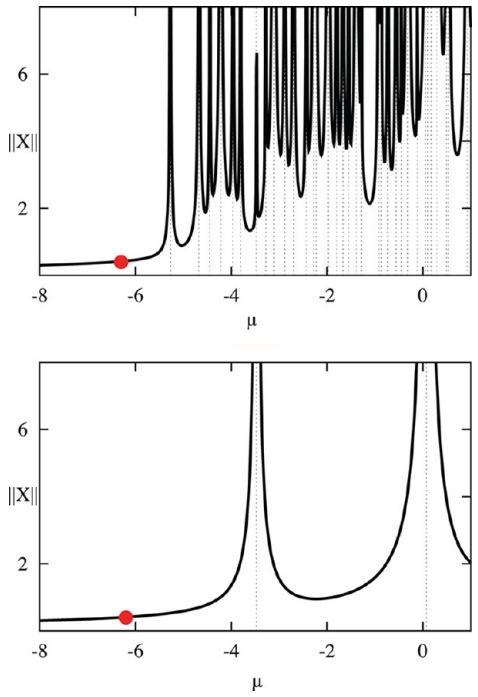

Figure 1. Step length function $\|X(\mu)\|$ plotted for a 100-dimensional model system on a range $-8 < \mu < 1$. Top: the step length function for a full space Hessian is displayed. Bottom: the step length function for a reduced space (6-dimensional) Hessian is displayed.

minimization step for a step length $\|X(\mu)\| = 0.4$ is marked with a bold red dot.

2.2. Level Shifted Newton Equations as an Eigenvalue Problem. The level-shifted Newton equation (eq 7) may be obtained by solving an augmented Hessian eigenvalue equation.[9,12,16]

$$\boldsymbol{A}(\alpha)\begin{pmatrix}1\\ \boldsymbol{x}(\alpha)\end{pmatrix} = \mu\begin{pmatrix}1\\ \boldsymbol{x}(\alpha)\end{pmatrix} \tag{13}$$

where $\boldsymbol{A}(\alpha)$ is the Hessian augmented with one extra dimension containing the gradient multiplied by a parameter α.

$$\boldsymbol{A}(\alpha) = \begin{pmatrix}0 & \alpha\boldsymbol{\xi}^{[1]^T}\\ \alpha\boldsymbol{\xi}^{[1]} & \boldsymbol{\xi}^{[2]}\end{pmatrix} \tag{14}$$

The parameter α is used to ensure that the step length is equal to h. The two components of eq 13 may be expressed as

$$\alpha\boldsymbol{\xi}^{[1]^T}\boldsymbol{x}(\alpha) = \mu \tag{15}$$

$$(\boldsymbol{\xi}^{[2]} - \mu\mathbf{1})\boldsymbol{x}(\alpha) = -\alpha\boldsymbol{\xi}^{[1]} \tag{16}$$

where the second component is the level shifted Newton equation of eq 7 with the solution

$$\boldsymbol{\kappa}(\mu) = \alpha^{-1}\boldsymbol{x}(\alpha) \tag{17}$$

The lowest eigenvalue of the augmented Hessian $\boldsymbol{A}(\alpha)$ is lower than the lowest eigenvalue of the Hessian $\boldsymbol{\xi}^{[2]}$, according to the Hylleraas−Undheim−McDonald theorem.[17,18] Thus, by choosing the lowest eigenvalue of eq 13 and by adjusting α to give $\|\alpha^{-1}\boldsymbol{x}(\alpha)\| = h$, we may generate a step of length h where $\mu < \varepsilon_1$ and which therefore determine the minimum on the boundary of the trust region.

2.3. Iterative Solution of the Level Shifted Newton Equation. The eigenvalue problem in eq 13 may be solved iteratively in a reduced space,[19] where—in each iteration—a vector is generated from the residual of eq 16 and added to the reduced space basis. In a reduced space basis, the solution to eq 13 may be computed repeatedly at an insignificant cost to determine the α which yields a solution fulfilling $\|\alpha^{-1}\boldsymbol{x}(\alpha)\| = \|\boldsymbol{\kappa}(\mu)\| = h$. Below we describe how the starting trial vectors are chosen and outline an arbitrary iteration n in the iterative procedure.

2.3.1. Start Trial Vectors. To separate the eigensolution into a component in the gradient direction $\boldsymbol{\xi}^{[1]}$ and a component orthogonal to this direction, as in eq 13, the structure of the reduced space vectors $\{\boldsymbol{B}_i\}$ is chosen as

$$\boldsymbol{B}_i = \begin{pmatrix}1\\ \mathbf{0}\end{pmatrix} \qquad \text{if } i = 0 \tag{18}$$

$$\boldsymbol{B}_i = \begin{pmatrix}0\\ \boldsymbol{b}_i\end{pmatrix} \qquad \text{if } i > 0 \tag{19}$$

where the first trial vector $\boldsymbol{b}_1$ is set equal to the normalized gradient

$$\boldsymbol{b}_1 = \|\boldsymbol{\xi}^{[1]}\|^{-1}\boldsymbol{\xi}^{[1]} \tag{20}$$

In addition to the normalized gradient, we include two extra vectors to the set of trial vectors before starting the iterative procedure. The trial vector $\boldsymbol{b}_2$ is obtained from the residual of eq 16

$$\boldsymbol{R} = -\boldsymbol{\xi}^{[1]} - (\boldsymbol{\xi}^{[2]} - \mu\mathbf{1})\alpha^{-1}\boldsymbol{x}(\alpha) \tag{21}$$

setting $\mu = 0$ and $\alpha^{-1}\boldsymbol{x}(\alpha) = \boldsymbol{b}_1 = \|\boldsymbol{\xi}^{[1]}\|^{-1}\boldsymbol{\xi}^{[1]}$. This gives

$$\boldsymbol{b}_2 = \boldsymbol{R}_1 = -\boldsymbol{\xi}^{[1]} - \|\boldsymbol{\xi}^{[1]}\|^{-1}\boldsymbol{\xi}^{[2]}\boldsymbol{\xi}^{[1]} \tag{22}$$

For small values of μ the $\boldsymbol{b}_2$ trial vector is a good first estimate of the solution vector. The third trial vector $\boldsymbol{b}_3$ is obtained identifying the lowest diagonal Hessian element $[\boldsymbol{\xi}^{[2]}]_{pq,pq}$, and use the vector that generates this diagonal element as trial vector.

$$\boldsymbol{b}_3 = \|\boldsymbol{V}\|^{-1}\boldsymbol{V}, \qquad V_{rs} = \delta_{pr}\delta_{qs} - \delta_{qr}\delta_{ps} \tag{23}$$

This trial vector is included because μ has to be lower than the lowest eigenvalue ε_1 which can be estimated from the lowest Hessian diagonal element and is lower than this, according to Hylleraas−Undheim−MacDonalds theorem.[17,18] Introducing the lowest Hessian diagonal element explicitly in the reduced space Hessian pushes μ toward the correct level shift range $\mu < \varepsilon_1$. We note that the exact diagonal Hessian elements $[\boldsymbol{\xi}^{[2]}]_{rs,rs}$ are calculated prior to the iterative scheme since they are needed for the preconditioning of new trial vectors as described in Section 2.3.2. The trial vectors are orthonormalized using the Gram−Schmidt procedure such that

$$\boldsymbol{b}_i^T\boldsymbol{b}_j = \delta_{ij} \tag{24}$$

When a Gram−Schmidt orthogonalization is used the first trial vector $\boldsymbol{b}_1$ always stays the normalized gradient as required to apply the intermediate normalization of eq 13 in the reduced space.

2.3.2. Iteration n. The first component of eq 13 is solved exactly in the reduced space implementations when the gradient direction is used as a trial vector (trial vector $\boldsymbol{b}_1$), as may be seen from eq 15. In the iterative algorithm we therefore effectively solve the second component of eq 13, that is, the level shifted Newton equation (eq 16).

Assume we have generated (including the starting vectors) n trial vectors $\boldsymbol{b}^n = \{\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_n\}$ during the iterative algorithm. The eigenvalue equation (eq 13) expressed in the set of trial vectors $\boldsymbol{B}^n = \{\boldsymbol{B}_0, \boldsymbol{B}_1, ..., \boldsymbol{B}_n\}$ is

$$\boldsymbol{A}(\alpha)^{B_n}\begin{pmatrix}1\\ \boldsymbol{x}^{b_n}(\alpha)\end{pmatrix} = \mu\begin{pmatrix}1\\ \boldsymbol{x}^{b_n}(\alpha)\end{pmatrix} \tag{25}$$

The elements of the augmented Hessian in the reduced space in eq 25, $\boldsymbol{A}(\alpha)^{B_n}$ is given by

$$\boldsymbol{A}(\alpha)^{B_n}_{00} = 0 \tag{26}$$

$$\boldsymbol{A}(\alpha)^{B_n}_{10} = \boldsymbol{A}(\alpha)^{B_n}_{01} = \alpha\boldsymbol{b}_1^T\boldsymbol{\xi}^{[1]} = \alpha\|\boldsymbol{\xi}^{[1]}\| \tag{27}$$

$$\boldsymbol{A}(\alpha)^{B_n}_{i0} = \boldsymbol{A}(\alpha)^{B_n}_{0i} = 0, \qquad i > 1 \tag{28}$$

$$\boldsymbol{A}(\alpha)^{B_n}_{ij} = [\boldsymbol{\xi}^{[2]}]^{b_n}_{ij} = \boldsymbol{b}_i^T\boldsymbol{\xi}^{[2]}\boldsymbol{b}_j \qquad i, j = 1, 2, \cdots, n \tag{29}$$

The reduced space augmented Hessian thus has the form

dx.doi.org/10.1021/ct300473g | J. Chem. Theory Comput. 2012, 8, 3137−3146

$$\mathbf{A}(\alpha)^{B_n} = \begin{pmatrix} 0 & \alpha\|\boldsymbol{\xi}^{[1]}\| & 0 & \cdots & 0 \\ \alpha\|\boldsymbol{\xi}^{[1]}\| & [\boldsymbol{\xi}^{[2]}]^{b_n}_{11} & [\boldsymbol{\xi}^{[2]}]^{b_n}_{12} & \cdots & [\boldsymbol{\xi}^{[2]}]^{b_n}_{2n} \\ 0 & [\boldsymbol{\xi}^{[2]}]^{b_n}_{21} & [\boldsymbol{\xi}^{[2]}]^{b_n}_{22} & \cdots & [\boldsymbol{\xi}^{[2]}]^{b_n}_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & [\boldsymbol{\xi}^{[2]}]^{b_n}_{n1} & [\boldsymbol{\xi}^{[2]}]^{b_n}_{n2} & \cdots & [\boldsymbol{\xi}^{[2]}]^{b_n}_{nn} \end{pmatrix} \tag{30}$$

The optimal solution to eq 16 in the reduced space is

$$\boldsymbol{x}(\alpha) = \sum_{i=1}^{n} x_i^{b^n}(\alpha)\boldsymbol{b}_i \qquad i = 1, 2, \cdots, n \tag{31}$$

Equation 25 is a matrix representation of eq 13 in the orthonormal basis $\boldsymbol{B}^n$. The first component of eq 25 is a matrix representation of eq 15 and becomes

$$\alpha\|\boldsymbol{\xi}^{[1]}\| x_1^{b^n}(\alpha) = \mu \tag{32}$$

which is solved exactly. Solving the eigenvalue equation thus means solving the level shifted Newton equation which is the second component of eq 13. In the reduced space the second component is a matrix representation of eq 16 in the orthonormal basis $\boldsymbol{b}^n$.

$$([\boldsymbol{\xi}^{[2]}]^{b^n} - \mu\boldsymbol{I})\boldsymbol{x}^{b^n}(\alpha) = -\alpha[\boldsymbol{\xi}^{[1]}]^{b^n} \tag{33}$$

where

$$[\boldsymbol{\xi}^{[1]}]^{b_n}_i = \alpha\boldsymbol{\xi}^{[1]^T}\boldsymbol{b}_i = \begin{cases} \alpha\|\boldsymbol{\xi}^{[1]}\| & i = 1 \\ 0 & i = 2, 3, \cdots, n \end{cases} \tag{34}$$

Our goal is to adapt α such that the solution for the lowest eigenvector of eq 25 in the reduced space satisfies

$$\|\alpha^{-1}\boldsymbol{x}^{b^n}(\alpha)\| = h \tag{35}$$

The small dimension of the reduced space makes it straightforward to determine this solution by repeated diagonalization in a bisectional search for the α satisfying eq 35. When the appropriate α has been determined, the trial vector $\boldsymbol{b}_{n+1}$ is obtained from the residual (eq 21) as

$$\boldsymbol{R}_n = -\boldsymbol{\xi}^{[1]} - (\boldsymbol{\xi}^{[2]} - \mu\mathbf{1})\alpha^{-1}\boldsymbol{x}^{b_n}(\alpha) \tag{36}$$

by inserting eq 31 as the solution. To reduce the condition number and thus improve convergence, a preconditioner is applied to each generated trial vector. We use the diagonal preconditioner $([\boldsymbol{\xi}^{[2]}]_{\mathrm{dia}}-\mu\mathbf{1})^{-1}$, where $[\boldsymbol{\xi}^{[2]}]_{\mathrm{dia}}$ is a matrix containing the precomputed full space diagonal Hessian elements on the diagonal.

$$\boldsymbol{R}_n^{\mathrm{P}} = ([\boldsymbol{\xi}^{[2]}]_{\mathrm{dia}} - \mu\mathbf{1})^{-1}\boldsymbol{R}_n \tag{37}$$

$\boldsymbol{R}_n^{\mathrm{P}}$ is orthonormalized against all previous trial vectors

$$\boldsymbol{b}_{n+1} = \hat{P}\boldsymbol{R}_n^{P} \tag{38}$$

where $\hat{P}$ refers to performing the Gram−Schmidt orthonormalization. $\boldsymbol{b}_{n+1}$ is added to the set of trial vectors, and iteration $n + 1$ may start by increasing the dimension of eq 25 by one. The iterative procedure is thus established.

The reduced space iterative procedure is terminated when the residual computed in eq 36 has been reduced by some factor k relative to the residual computed using the three start vectors $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, and $\boldsymbol{b}_3$. For example, if $\boldsymbol{R}_n < k\boldsymbol{R}_3$ we exit, and the step used to transform the orbitals according to eq 1 is

$$\boldsymbol{\kappa}(\mu) = \alpha^{-1}\sum_{i=1}^{n} x_i^{b^n}(\alpha)\boldsymbol{b}_i \tag{39}$$

which is the reduced space representation of the step in eq 17. The iterations which are used for solving the level shifted Newton equation are termed micro iterations while the iterations for the trust region minimization are termed macro iterations. The macro iterations are converged when the gradient norm $\|\boldsymbol{\xi}^{[1]}\|$ is below some threshold. The numerical values of the convergence factor k and the gradient norm threshold will be discussed in Section 3.1.

2.4. Conservative Estimate of Trust Region Followed by a Line Search. The step length norm $\|\boldsymbol{\kappa}\|$ is a size extensive quantity. For large molecular systems $\|\boldsymbol{\kappa}\|$ is therefore not a good measure to use for the total step lengths as a large $\|\boldsymbol{\kappa}\|$ may be due to a sum of many small elements which leads to a fast converging exponential matrix, or it may be due to one single large element leading to steps far outside the region where the second order Taylor expansion can be trusted. Salek et al.[11] suggested to use max $|\kappa_{pq}|$ as a size intensive measure. However, max $|\kappa_{pq}|$ is not invariant with respect to a basis transformation making it nontractable to use as a total step size measure.

For both small and large molecular systems there exists a region of ξ where ξ is approximated well by its second order Taylor expansion $\Lambda(\boldsymbol{\kappa})$ of eq 3. Instead of updating the trust region in a self-adaptive manner as in a standard trust region minimization, we suggest an alternative approach. In this approach we use a conservative trust radius in all iterations without update to determine the direction of the step and carry out a line search along the obtained step direction to determine the length of the step. The conservative trust radius h^{c} has to be small to ensure that the linear term in the Taylor expansion of the exponential function is dominating (h^{c} less than $\pi/4$).[9] However the step cannot be too small as the step (see eq 16) has to be dominated by the Hessian contribution and not by the level shift contribution as the level shift contribution represents a step in the steepest descent direction which is uphill directed for directions corresponding to negative Hessian eigenvalues. When function values are easily evaluated as for the localization functions it is extremely effective to carry out a line search along the direction of a conservative step size.

2.4.1. Illustrative Example. To illustrate how steps are determined in the trust region minimization algorithm we consider the 100-dimensional model system in Figure 1 (top). Assume we have carried out an iterative procedure and determined a six dimensional reduced space with the step length function given in Figure 1 (bottom) where the step length function is plotted in the interval $-8 < \mu < 1$. The minimizing step in the full 100-dimensional model space must be in the region $\mu < -5.2$ whereas for the six-dimensional model space the level shift can only be restricted to $\mu < -3.7$, where -5.2 and -3.7 is the lowest Hessian eigenvalue in the full and reduced space respectively. If μ is chosen in the region between -5.2 and -3.7 the step in the reduced space when expanded to the full space will have components that are uphill directed. Choosing a conservative trust radius $h^{\mathrm{c}} = 0.4$ gives a step which in Figure 1 (bottom) is marked with a bold dot at $\mu = -6.2$. The step for $\mu < -6.2$ thus will be downhill directed as $\mu = -6.2$ is well below the lowest eigenvalue of -5.2 for the full Hessian.

For a conservative trust radius, steps are thus taken where all the Hessian eigenvector components in the full space are

downhill directed. The weight of the different eigenvector components differ slightly depending on the value that is actually associated to the conservative trust radius, but when a line search is carried out afterward approximately the same total step is obtained. For example, for conservative trust radii h^c = 0.4 and h^c = 0.2 where both steps are followed by a line search, about the same total step will be obtained.

3. RESULTS

In this section we give examples where the trust region method is used to localize a set of occupied and virtual HF orbitals. We consider localization based on minimizing powers of the orbital variance[7]

$$\xi_m = \sum_i \left(\sum_x \langle il(x - \langle il|x|li\rangle)^2|li\rangle \right)^m = \sum_i \Omega_i^m \qquad (40)$$

where x is a Cartesian coordinate and m is a positive integer. The minimization of other localization functions have been found to proceed in much the same way as for ξ_m. Increasing m in eq 40 increase the penalty for having orbitals with a large orbital variance. For m = 1 eq 40 reduce to the Boys localization function. As shown by Jansik et al.[7] m = 2 is required to remove outliers, that is, single orbitals with a large variance. Convergence results will be presented for both ξ_1 and ξ_2. Since ξ_2 is more demanding than ξ_1, the focus will be on ξ_2. The locality of the localized sets of orbitals will be quantified by the orbital spread, $(\Omega_i)^{1/2}$, of the least local orbital in the set.

Orbital localization convergence results will be presented for insulin ($C_{257}N_{65}O_{77}S_6H_{382}^-$) in a cc-pVDZ basis (7604 basis functions), C_{240} in a cc-pVDZ basis (3360 basis functions), and for C_{240} in a cc-pVTZ basis (7200 basis functions). The first system is chosen because of sheer size, proving the capacity and power of the trust region procedure. The latter system is chosen because it is traditionally deemed to be a delocalized system, and we want to show that orbitals for such systems can successfully be localized when using the trust region optimization. Convergence results will be presented for core, valence, and virtual orbitals to illustrate the performance of the trust region minimization. The core and valence orbitals are localized separately to maintain a physical separation of these

Table 1. Summary of the Settings Used for the Results Presented in This Paper

specification	value
start trust radius, h^c	0.4
macro threshold	10^{-5}
red. factor, global (k_g)	0.05
red. factor, local (k_l)	0.005
no. of micro iterations	50

Table 2. Orbital Spreads of the Least Local Orbital (Orbital with Largest Orbital Spread) in the Sets of Core ($\max[\sigma_p]_{core}$), Valence ($\max[\sigma_p]_{val}$), and Virtual ($\max[\sigma_p]_{virt}$) Orbitals[a]

molecule/basis	starting guess	m	$\max[\sigma_p]_{core}$	$\max[\sigma_p]_{val}$	$\max[\sigma_p]_{virt}$
insulin/cc-pVDZ	CMO	1	0.44 au	2.21 au	3.56 au
insulin/cc-pVDZ	LCM	1	0.44 au	2.17 au	3.62 au
insulin/cc-pVDZ	CMO	2	0.44 au	2.13 au	2.70 au
insulin/cc-pVDZ	LCM	2	0.44 au	2.12 au	2.67 au
C_{240}/cc-pVDZ	CMO	2	0.31 au	2.51 au	2.47 au
C_{240}/cc-pVTZ	CMO	2	0.31 au	2.51 au	2.55 au

[a]The orbitals are obtained starting from either CMO or LCM orbitals using ξ_m.

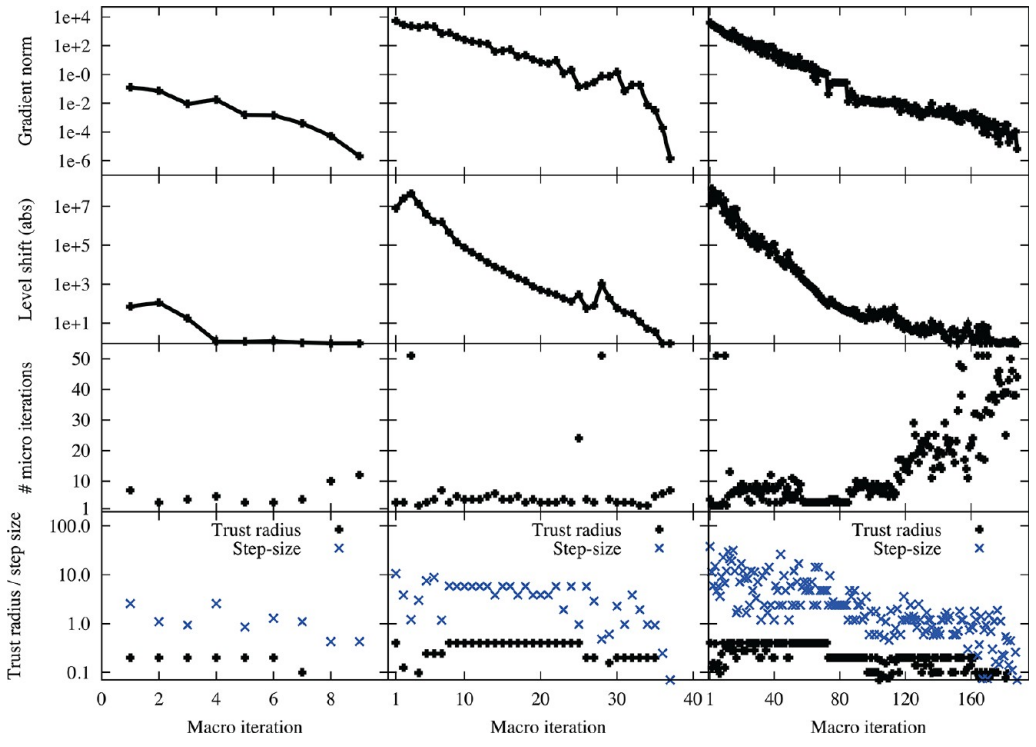

Figure 2. Insulin cc-pVDZ orbital localization convergence rates for the 429 core (left), 1117 valence (middle), and 6058 virtual (right) orbitals minimizing ξ_2 starting from CMOs.

dx.doi.org/10.1021/ct300473g | J. Chem. Theory Comput. 2012, 8, 3137−3146

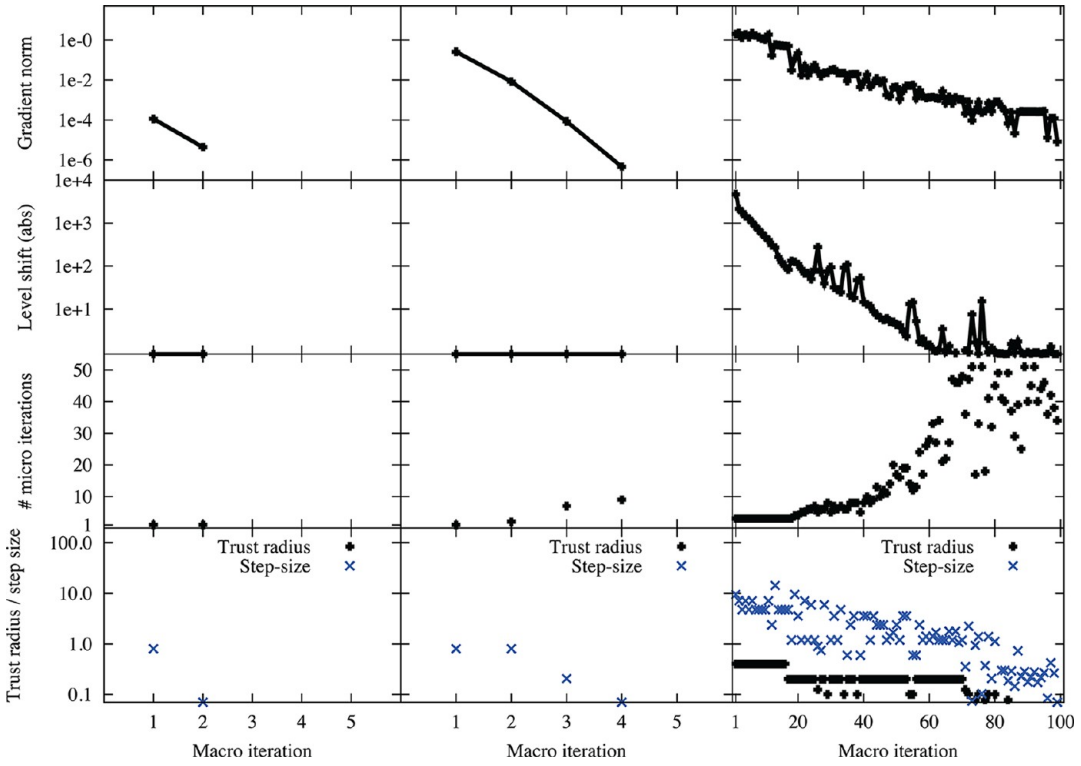

Figure 3. Insulin cc-pVDZ orbital localization convergence rates for the 429 core (left), 1117 valence (middle), and 6058 virtual (right) orbitals minimizing ξ_2 starting from LCM orbitals.

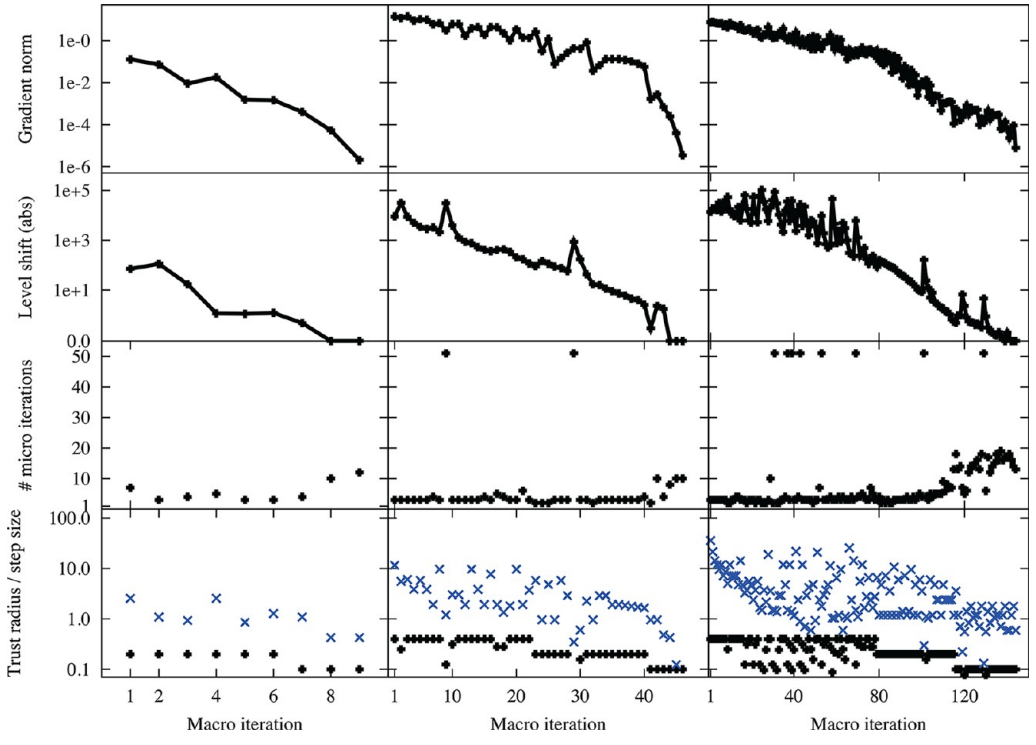

Figure 4. Insulin cc-pVDZ orbital localization convergence rates for the 429 core (left), 1117 valence (middle), and 6058 virtual (right) orbitals minimizing ξ_1 starting from CMOs.

two sets of orbitals which will become lost if they are simultaneously localized.

We present rates of convergence for localizations starting from both canonical molecular orbitals (CMOs) and least change molecular (LCM) HF orbitals.[20] The LCM HF orbitals are obtained using the three level optimization strategy[21] where small orbital transformations are carried out at each step of the optimization to preserve the locality of the atomic orbitals basis best possible.[22] The CMOs constitute a completely delocalized set of starting orbitals, whereas the LCM orbitals are semilocal.

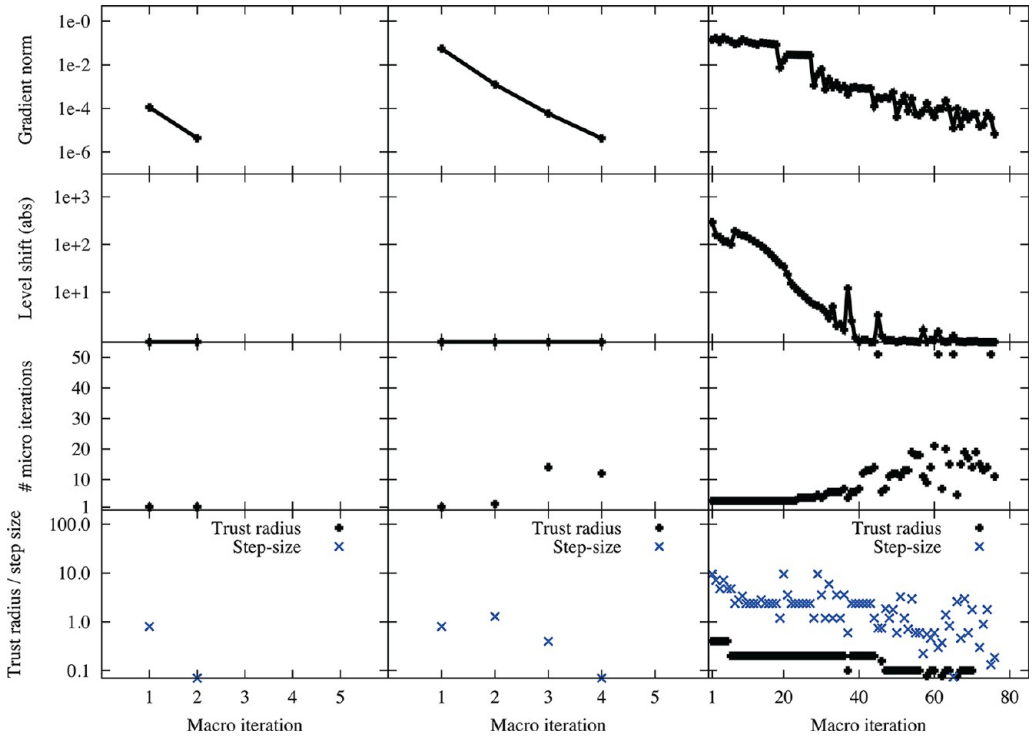

Figure 5. Insulin cc-pVDZ orbital localization convergence rates for the 429 core (left), 1117 valence (middle), and 6058 virtual (right) orbitals minimizing ξ_1 starting from LCM orbitals.

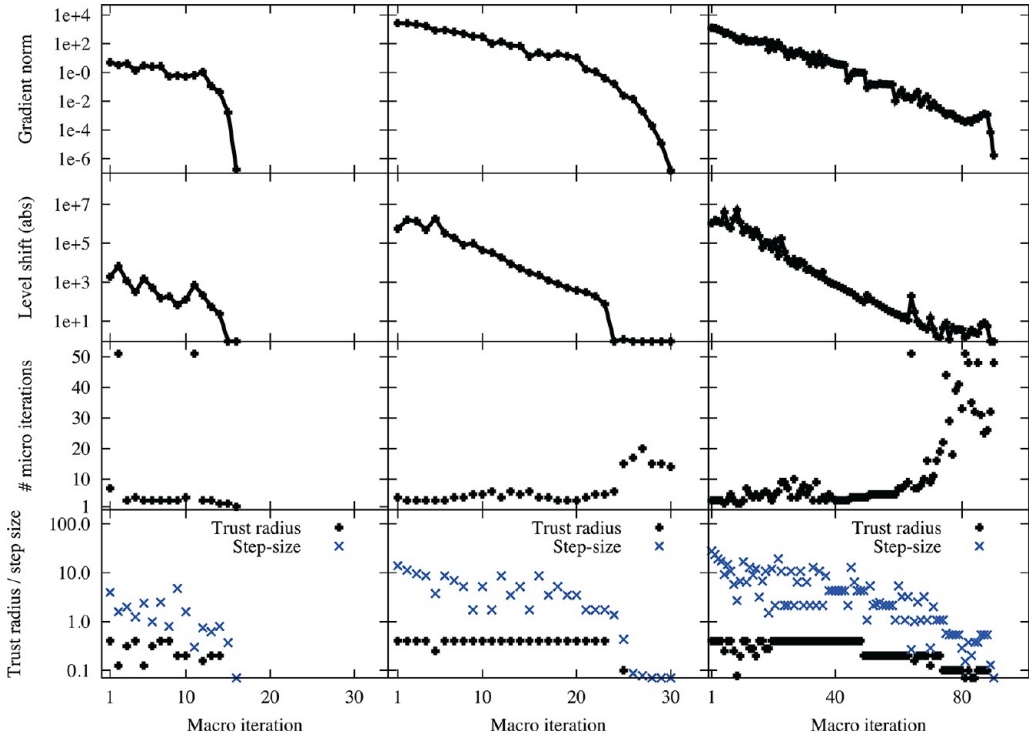

Figure 6. C_{240} cc-pVDZ orbital localization convergence rates for the 240 core (left), 480 valence (middle), and 2640 virtual (right) orbitals minimizing ξ_2 starting from CMOs.

Because of being semilocal, LCM starting orbitals lead to faster, more uncomplicated convergence than the CMOs, and also lessen the risk of being trapped in local minima, for example, in symmetry solutions corresponding to nonlocal orbitals. This will be discussed in detail in ref 22.

3.1. Algorithmic Details about the Trust Region Minimization. We now describe details for the trust region minimization of the localization functions. The sequence of trust region iterations for minimizing the localization function is termed macro iterations. Convergence is obtained when the gradient norm of the localization function is smaller than 10^{-5}.

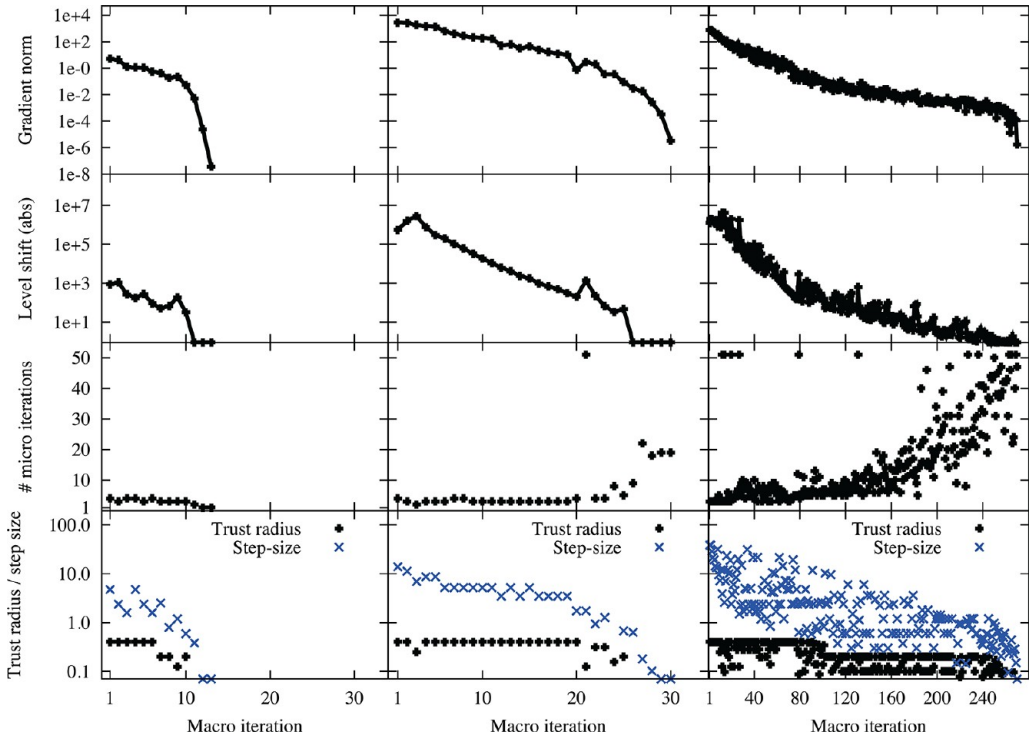

Figure 7. C_{240} cc-pVTZ orbital localization convergence rates for the 240 core (left), 480 valence (middle), and 6480 virtual (right) orbitals minimizing ξ_2 starting from CMOs.

The sequence of iterations for solving the level shifted Newton equations are termed micro iterations, and the maximum number of micro iterations is set to 50. Using preconditioning as described in Section 2.2 and using a conservative step length (0.4), 50 iterations are generally more than sufficient to solve the level shifted Newton equation to the preset tolerance threshold both in the global and the local region. In some macro iterations it is difficult to converge the level shifted Newton equation. This signifies that the level shift has been chosen in the wrong region such that the level shifted Hessian $(\boldsymbol{\xi}^{[2]}-\mu\mathbf{1})$ is not positive definite, and the preconditioning of the step in eq 37 therefore is ill defined.[23] To make $(\boldsymbol{\xi}^{[2]}-\mu\mathbf{1})$ positive definite a larger level shift has to be chosen, and this is done by reducing the size of the trust radius h^c. In practice we have found that if the residual is not decreased by a factor 0.1 (relative to residual computed using the starting vectors described in Section 2.3.1) in five successive micro iterations we have to reduce the trust radius by a factor 1/2 to increase the numerical size of the level shift. In a few cases the step reduction has been performed more than one time.

The settings are listed in Table 1. A relative threshold k has been used when converging the micro iterations. k_g is used in the global region, where we are satisfied with a step in a reasonably correct downhill direction. As we enter the local region a tighter convergence factor k_l is used, to ensure getting the correct step toward the minimum. Quadratic convergence in the local region requires the Newton equations are solved to a residual threshold that is equal to the squared norm of the gradient. Since the local region is a very small fraction of the total minimization a relative threshold $k_l = 0.005$ has been used without compromising the efficiency of the algorithm. All results presented in this paper are obtained using the algorithm settings in Table 1.

3.2. ξ_2 Localization for Insulin from CMOs. In Figure 2 we have for insulin in a cc-pVDZ basis reported the convergence characteristics for localizing the core, valence, and virtual HF orbitals starting with a set of canonical HF (CMO) orbitals. For each macro iteration, we display the gradient norm $\|\boldsymbol{\xi}_2^{[1]}\|$, the level shift μ, and the number of micro iterations used for solving the level shifted Newton equations. In addition we report (for $\mu \neq 0$) the conservative trust radius for the level shifted Newton equations and the norm of the step from the line search. For $\mu = 0$ only the step norm is reported.

The convergence of the macro iterations is nearly monotonic for all localizations (core, valence, and virtual) but with some small bumps. The core orbitals are by far the simplest to localize; $\|\boldsymbol{\xi}_2^{[1]}\|$ starts out with a norm of about 10^{-1} and the threshold 10^{-5} is reached in 9 iterations. In the initial three iterations level shifts are required . The line search is enlarging the step sizes obtained solving the level shifted Newton equations to about 2.6. Note that 0.2 has been used as a conservative trust radius for the core localization.

The valence and virtual orbitals are much more difficult to localize than the core orbitals with the virtual being the most difficult. The valence orbitals starts out with a $\|\boldsymbol{\xi}_2^{[1]}\|$ of about 10^4 and convergence is obtained in 37 iterations to a gradient norm 10^{-5}. The level shifts in the initial iterations are around -10^7 signifying a complex structure of the localization function with negative Hessian eigenvalues of the same order of magnitude as the level shift. The numerical value of the level shifts are gradually decreased during the iterative procedure, and at convergence the Hessian is positive definite.

The level shifted Newton equations are generally very simple to solve and requires less than 10 micro iterations for simultaneously determining the level shift and converging the Newton equations to the preset relative threshold. In the initial and last iterations step reduction was imposed on the trust

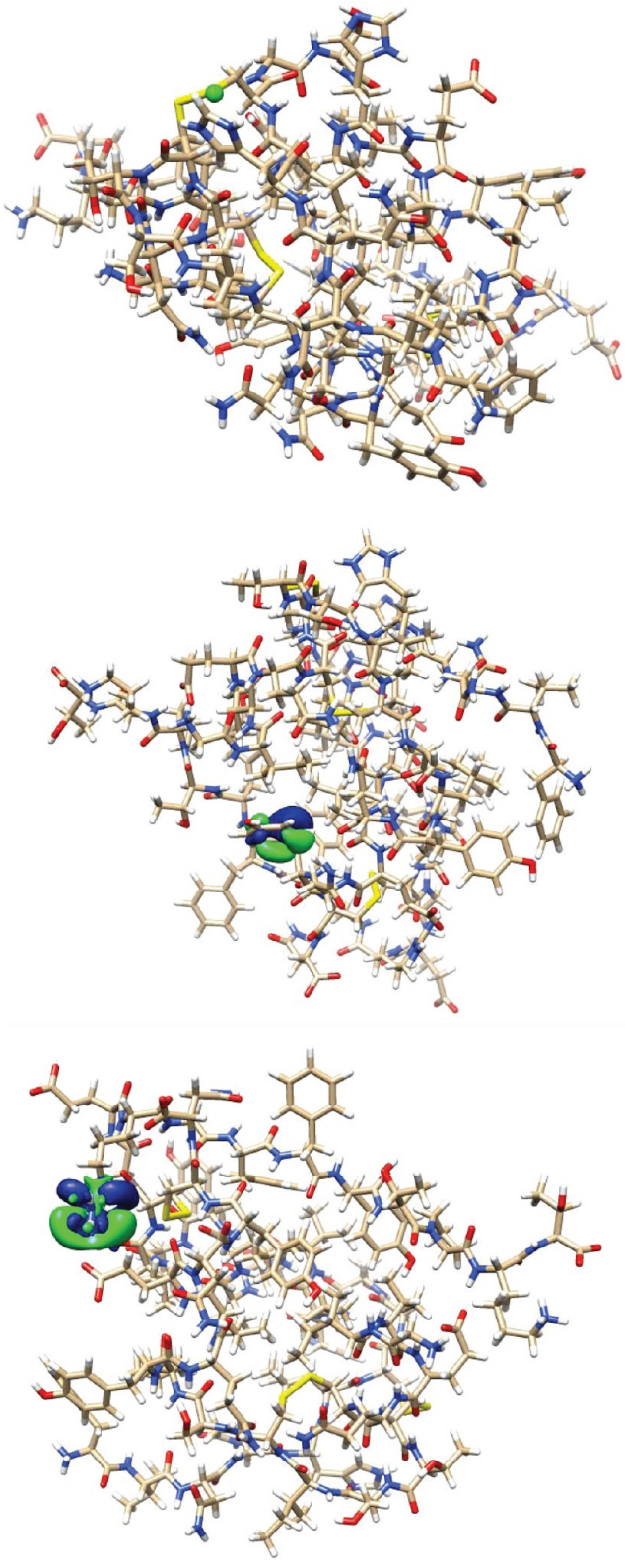

Figure 8. Core (top), least local valence (middle), and least local virtual (bottom) for insulin using ξ_1 starting from CMOs plotted using a contour value of 0.03. The orbital spreads are given in the first row of Table 2.

radius to converge the equations. The line search along the level shifted Newton directions increase the step sizes significantly. In some iterations the step size is actually increased to about 10. The line search is thus extremely important for an efficient minimization as very large steps can then safely be taken. In contrast, such large steps are not allowed using a standard trust region update since the Taylor expansion is only uniquely defined for one period of the exponential function. In a few macro iterations the maximum number of micro iterations is reached and reduction of the trust radius have been imposed several times. This indicates that we have reached a very complicated region for the optimization function. However, the algorithm is able to recover and proceeds toward the minimum.

The localization of the virtual orbitals is a little more difficult than the localization of the valence orbitals requiring a total of about 190 iterations to converge from a gradient norm of about 10^4 to 10^{-5}. The major difference between the valence and virtual orbital localization is that the step reductions of the trust radius are encountered more often and that the line search in general leads to larger steps. Further, in the last part of the macro iterations more trial vectors are required for solving the Newton equations. In this region the local convergence threshold norm 0.005 is also imposed. The overall convergence appear to be equally stable for virtual and for valence orbitals.

The orbital spreads for the least local core, valence, and virtual orbitals obtained from the localization are listed in row 3 of Table 2.

3.3. ξ_2 Localization for Insulin from LCM Orbitals. In Figure 3 we report the convergence for an insulin calculation similar to the one in Section 3.2 except that a set of LCM orbitals are used as starting guess.[20,22] As seen from Figure 3 LCM orbitals give an extremely good starting guess for the core and the valence orbitals. The core orbitals converge in two macro iterations and the valence orbitals in four iterations, and no level shifts are required. For the virtual orbitals the starting guess is also greatly improved compared to a CMO starting guess. The virtual orbital localization converges in 100 macro iterations starting with a gradient norm of about 1. Comparing the convergence of the virtual orbitals in Figure 2 with the one in Figure 3 we see that the one in Figure 3 is similar to the last part of the one in Figure 2 (after the gradient norm has dropped below 1). All features in the two graphs are similar; the reduction in level shifts, the trust radius, the number of micro iterations, and the step sizes obtained from the line search. The orbital spreads for the least local core, valence, and virtual orbitals obtained from the localization are listed in row 4 of Table 2.

3.4. Other Examples of Localization. In Figure 4 and Figure 5 we report convergence characteristic for a cc-pVDZ insulin calculation using ξ_1 (Boys localization function) with a CMO starting guess in Figure 4 and a LCM starting guess in Figure 5. The only difference between the results reported in Figure 4 and Figure 5 and in Figure 2 and Figure 3 is that the ξ_1 localization function is used in Figure 4 and Figure 5 while ξ_2 is used in Figure 2 and Figure 3. The overall convergence characteristics are very similar in Figure 2 and Figure 4 where CMO starting orbitals are used and in Figure 3 and Figure 5 where LCM starting orbitals are used. LCM orbitals are a much better set of starting orbitals than CMO orbitals. For the core and valence orbitals very few iterations are required to obtain a set of localized orbitals while for the virtual orbitals the number of iterations is about halved. The obtained orbitals are less local than the orbitals obtained using ξ_2 as may be seen by comparing the first four rows of Table 2. Nonetheless, the orbitals from the ξ_1 localization are local, as illustrated in Figure

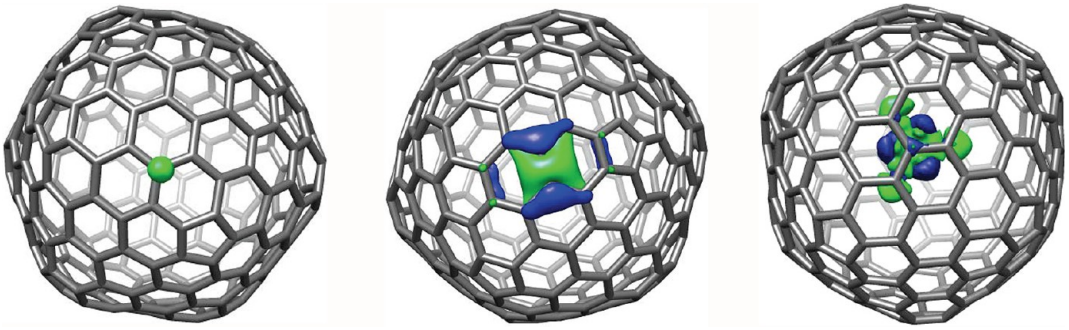

Figure 9. Core (left), least local valence (middle), and least local virtual (right) for C_{240} in a cc-pVTZ basis localized using ξ_2 plotted using a contour value of 0.03. The orbital spreads are given in the last row of Table 2.

8 where the core, valence, and virtual orbitals of row 1 in Table 2 are plotted using a contour value of 0.03.

To emphasize the efficiency of the trust region localization algorithm we have in Figure 6 and Figure 7 respectively for ξ_2 reported the convergence characteristics for a C_{240} calculation using a cc-pVDZ and cc-pVTZ basis. In both cases a CMO starting set of orbitals was used. As clearly seen from Figure 6 and Figure 7 no difficulties are encountered localizing the orbitals for C_{240} even though this system is expected to be highly delocal. The locality of the orbitals may be seen from Table 2 (row 5 and 6) and from Figure 9 where the least local core, valence, and virtual orbital of row 6, Table 2 are plotted using a contour value of 0.03.

4. SUMMARY

In this paper we have described how the trust region method may successfully be used to obtain local HF orbitals both for the set of core, valence, and virtual orbitals. The trust region method and its implementation is described in detail. A variant of the standard trust region method is proposed where the trust region update is replaced with an algorithm where a step direction is obtained based on a conservative estimate of the trust region followed by a line search along the obtained direction to obtain the actual step length.

Using the presented trust region minimization, convergence is straightforwardly obtained for all considered molecules and localization functions. In contrast, the standard approach of a Jacobi sweep of iterations has in general failed to give a set of local virtual orbitals. The use of the trust region algorithm has made it a simple and straightforward task to determine local orbitals both for core, valence, and virtual HF orbitals.

AUTHOR INFORMATION

Corresponding Author

*E-mail: idamh@chem.au.dk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the Lundbeck Foundation, the Danish Center for Scientific Computing (DCSC). The research leading to these results have received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 291371.

REFERENCES

(1) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 296.
(2) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.
(3) Boys, S. F. In *Quantum Theory of Atoms, Molecules and Solid State*; Löwdin, P.-O., Ed.; Academic: New York, 1966; p 253.
(4) Edmiston, C.; Ruedenberg, K. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
(5) Edmiston, C.; Ruedenberg, K. *J. Chem. Phys.* **1965**, *43*, S97–S116.
(6) Pipek, J.; Mezey, P. G. *J. Chem. Phys.* **1989**, *90*, 4916.
(7) Jansík, B.; Høst, S.; Kristensen, K.; Jørgensen, P. *J. Chem. Phys.* **2011**, *134*, 194104.
(8) Fletcher, R. *Practical Methods of Optimization*; Wiley: Chichester, U.K., 1980; Vol. *1*; pp 77–90.
(9) Aa. Jensen, H. J.; Jørgensen, P. *J. Chem. Phys.* **1984**, *80*, 1204–1211.
(10) Jørgensen, P.; Swanstrom, P.; Yeager, D. *J. Chem. Phys.* **1983**, *78*, 347–356.
(11) Salek, P.; Thøgersen, L.; Jørgensen, P.; Manninen, P.; Olsen, J.; Jansik, B.; Reine, S.; Pawlowski, F.; Tellgren, E.; Helgaker, T.; Coriani, S. *J. Chem. Phys.* **2007**, *126*, 114110.
(12) Lengsfield, B., III *J. Chem. Phys.* **1980**, *73*, 382.
(13) Cerjan, C.; Miller, W. *J. Chem. Phys.* **1981**, *75*, 2800–2806.
(14) Simons, J.; Jørgensen, P.; Taylor, H.; Ozment, J. *J. Phys. Chem.* **1983**, *87*, 2745–2753.
(15) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*, 1st ed.; Wiley: New York, 2000; p 93.
(16) Shepard, R.; Shavitt, I.; Simons, J. *J. Chem. Phys.* **1982**, *76*, 543.
(17) Hylleraas, E. A.; Undheim, B. *Z. Phys.* **1930**, *65*, 759.
(18) MacDonald, J. K. L. *Phys. Rev.* **1933**, *43*, 830.
(19) Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87.
(20) Ziółkowski, M.; Jansík, B.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2009**, *131*, 124112.
(21) Jansík, B.; Høst, S.; Johansson, M. P.; Olsen, J.; Jørgensen, P.; Helgaker, T. *Chem. Phys. Phys. Chem* **2009**, *11*, 5805.
(22) Høyvik, I.-M.; Jansik, B.; Kristensen, K.; Jørgensen, P. Three level optimization. **2012**, in preparation.
(23) Ziółkowski, M.; Weijo, V.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **2008**, *128*, 9606.