# Curvy steps for density matrix based energy minimization: tensor formulation and toy applications

MARTIN HEAD-GORDON [a] , YIHAN SHAO [a] , CHANDRA SARAVANAN [a] & CHRISTOPHER A. WHITE [b]

[a] Department of Chemistry , University of California, Berkeley and Chemical Sciences Division, Lawrence Berkeley National Laboratory , Berkeley, CA, 94720, USA

[b] Electro-Optic Devices Research, Agere Systems , 600 Mountain Ave, Murray Hill, NJ, 07974, USA
Published online: 25 Nov 2009.

PLEASE SCROLL DOWN FOR ARTICLE

# Curvy steps for density matrix based energy minimization: tensor formulation and toy applications

MARTIN HEAD-GORDON*[1], YIHAN SHAO[1], CHANDRA SARAVANAN[1]
and CHRISTOPHER A. WHITE[2]

[1] Department of Chemistry, University of California, Berkeley and Chemical
Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720,
USA
[2] Electro-Optic Devices Research, Agere Systems, 600 Mountain Ave, Murray Hill,
NJ 07974, USA

A unitary transformation-based framework is explained for varying the one-particle density matrix in a non-orthogonal expansion basis while preserving both the idempotency and electron number constraints. The equations are presented in a compact tensor formulation that permits the use of any representation for the matrices. A connection with the LNV method is established. Curvy steps, which are high order line searches along a chosen descent direction, are defined. Calculations with two toy model systems are presented to illustrate the nature of the displacements in this approach.

## 1. Introduction

The problem of minimizing the energy in a tight-binding or self-consistent field theory within a finite basis has traditionally been solved by diagonalization [1], which has cubic scaling computational complexity. With the advent of linear scaling techniques for the formation of the effective Hamiltonian matrix [2, 3] together with accurate tight-binding potentials, this has become the primary bottleneck in large electronic structure calculations. Much effort has been devoted to linear scaling algorithms that bypass the diagonalization bottleneck [4]. Most such methods have solved either for orbitals that are localized by some constraint or, alternatively, have solved for the density matrix using a modified Lagrangian whose stationary points match those of the true energy.

Recently an alternative was proposed based on utilizing unitary transformations of an initial idempotent density matrix [5]. Unitary transformations based on an exponential ansatz have been used occasionally in the molecular orbital basis [6, 7], and the new recognition was that they could also be applied in the atomic orbital basis. This could be potentially valuable as an alternative linear scaling algorithm, because the density matrix and Hamiltonian matrix are both sparse in this representation [8].

The purpose of this paper is to explore further the unitary transformation based approach. We shall first present the defining theory in a tensor formalism that applies to any coordinate system, including as particular cases the traditional molecular orbital basis and the atomic orbital basis. With this formalism, it is clear also how to construct displacements that are properly invariant to the choice of coordinates, as is desirable for a well behaved numerical method.

## 2. Theory

### 2.1. Unitary transformations in non-orthogonal representation

A non-orthogonal basis is naturally handled using tensor algebra [9–11]. Subscripts are covariant indices in the given atomic orbital basis, $\{|\phi_\mu\rangle\}$. Superscripts refer to the biorthogonal contravariant basis, which satisfies $\langle \phi^\mu | \phi_\nu \rangle = \delta^\mu_{\cdot\nu}$. Covariant indices are converted to contravariant by application of the inverse overlap matrix, $g^{\mu\nu} \equiv (S^{-1})_{\mu\nu}$, while the contravariant to covariant conversion is done with $S_{\mu\nu}$. If the same symbol appears as both a covariant and contravariant index, then summation is implied (Einstein convention).

Fundamentally, matrix elements of the Hamiltonian (one- and two-electron integrals) are expressed in the covariant representation (e.g., $F_{\mu\nu} \equiv \langle \phi_\mu | \hat{F} | \phi_\nu \rangle$) while the density matrix elements to be solved for are contravariant, $P^{\mu\nu} \equiv \langle \phi^\mu | \hat{\rho} | \phi^\nu \rangle$. We call this the 'covariant integral representation'. For real basis

* Author for correspondence. e-mail: mhg@cchem.
berkeley.edu

functions, Hermitian and anti-Hermitian matrices are symmetric and antisymmetric in this representation. However, equations isomorphic to operator equations are obtained only in the 'natural representation' where all kets are covariant and all bras are contravariant (thus the density and Fock matrices appear as $P^\mu_{\bullet\nu}$ and $F^\mu_{\bullet\nu}$, which are unsymmetric matrices). Often we derive equations in the natural representation and then transform to the covariant integral representation if desired at the end.

The adjoint operator (†) can be defined as follows in the natural representation: operators map into their adjoint operators, kets map into bras with covariant subscripts turned into contravariant superscripts, and the order of operations is reversed. Thus, e.g., $|\phi_\mu\rangle^\dagger = \langle\phi^\mu|$ and $\langle\phi^\mu|\hat{F}|\phi_\nu\rangle^\dagger = \langle\phi^\nu|\hat{F}^\dagger|\phi_\mu\rangle$. The second example shows that matrix elements do not map into their complex conjugates because of the mapping between kets and their biorthogonal bras that we have used to define the adjoint operator.

Following standard practice, a unitary operator $\hat{U}$ may be parametrized as the exponential of an anti-Hermitian operator $\hat{\Delta}$. In the natural representation these operators will be represented by tensors $U^\mu_{\bullet\nu}$ and $\Delta^\alpha_{\bullet\nu}$, which are related by

$$U^\mu_{\bullet\nu} = (e^\Delta)^\mu_{\bullet\nu} = 1 + \Delta^\mu_{\bullet\nu} + \frac{1}{2!}\Delta^\mu_{\bullet\lambda}\Delta^\lambda_{\bullet\nu} + \cdots \quad (1)$$

For brevity, we shall sometimes write $U^\mu_{\bullet\nu} = e^{\Delta^\mu_{\bullet\nu}}$, with the understanding that this means equation (1). The adjoint of the unitary transformation is represented by the following tensor:

$$(\hat{U}^\dagger)^\mu_{\bullet\nu} = (e^{\Delta^\dagger})^\mu_{\bullet\nu} = (e^{-\Delta})^\mu_{\bullet\nu} = 1 - \Delta^\mu_{\bullet\nu} + \frac{1}{2!}\Delta^\mu_{\bullet\lambda}\Delta^\lambda_{\bullet\nu} - \cdots \quad (2)$$

Likewise we shall abbreviate equation (2) as $(U^\dagger)^\mu_{\bullet\nu} = e^{-\Delta^\mu_{\bullet\nu}}$. By construction, $U^\mu_{\bullet\nu}$ satisfies

$$(U^\dagger U)^\mu_{\bullet\nu} = e^{-\Delta^\mu_{\bullet\alpha}}e^{\Delta^\alpha_{\bullet\nu}} = \delta^\mu_{\bullet\nu}. \quad (3)$$

Using the mixed tensor representation in equation (3) shows that, in a non-orthogonal basis, the unitary transformation is being used to preserve biorthogonality, rather than orthogonality.

We may write analogue of equation (3) in other representations, by transforming the indices appropriately, as in the two examples:

$$(U^\dagger)^{\bullet\lambda}_\mu S_{\lambda\sigma}U^\sigma_{\bullet\nu} = S_{\mu\nu}, \quad (4)$$

$$(U^\dagger)^\mu_{\bullet\lambda}S^{\lambda\sigma}U^{\bullet\nu}_\sigma = S^{\mu\nu}. \quad (5)$$

These relations provide a way of transforming fully covariant (or contravariant) matrices such that the metric (or inverse metric) is preserved. Furthermore,

the generators of the unitary rotation are most conveniently written in the covariant integral representation, via $\Delta^\mu_{\bullet\nu} = \Delta^{\mu\lambda}S_{\lambda\nu}$. $\Delta^{\mu\nu}$ must be antisymmetric, but $\Delta^\mu_{\bullet\nu}$ does not have this property.

## 2.2. Transformations of the density matrix that preserve idempotency and electron number

A new idempotent density matrix $\tilde{P}$ may be obtained by a unitary transform of another idempotent matrix. In the natural representation, we have

$$\tilde{P}^\mu_{\bullet\nu} = (U^\dagger)^\mu_{\bullet\lambda}P^\lambda_{\bullet\sigma}U^\sigma_{\bullet\nu} = e^{-\Delta^\mu_{\bullet\lambda}}P^\lambda_{\bullet\sigma}e^{\Delta^\sigma_{\bullet\nu}}. \quad (6)$$

This transformation preserves idempotency of the density matrix, $\tilde{P}^\mu_{\bullet\lambda}\tilde{P}^\lambda_{\bullet\nu} = \tilde{P}^\mu_{\bullet\nu}$, and electron number, $\tilde{P}^\mu_{\bullet\mu} = n$, if the initial density matrix $P$ has these properties. Alternatively, we may immediately transform to the covariant integral representation (compare equation (5)) to obtain

$$(U^\dagger)^\mu_{\bullet\lambda}P^{\lambda\sigma}U^{\bullet\nu}_\sigma = e^{-\Delta^\mu_{\bullet\lambda}}P^{\lambda\sigma}e^{\Delta^{\bullet\nu}_\sigma} = \tilde{P}^{\mu\nu}, \quad (7)$$

where $\Delta^\mu_{\bullet\nu} = \Delta^{\mu\lambda}S_{\lambda\nu}$ and $\Delta^{\bullet\nu}_{\bullet\nu} = S_{\mu\lambda}\Delta^{\lambda\nu}$. Equation (7) preserves idempotency, $\tilde{P}^{\mu\lambda}S_{\lambda\sigma}\tilde{P}^{\sigma\nu} = \tilde{P}^{\mu\nu}$, and electron number, $\tilde{P}^{\mu\nu}S_{\nu\mu} = n$, in this alternative representation.

These transformations of the density matrix may be evaluated by diagonalization of $\Delta$ followed by exponentiating its eigenvalues, and then pre- and post-multiplying by the eigenvectors [7]. However, this will involve inherently cubic computational effort. Instead we shall expand the transformations in powers of $\Delta$, which potentially will be useful in computational applications. In an orthogonal basis, the Baker–Hausdorff expansion compactly expresses the similarity transformation of a matrix with a matrix exponential as a series of commutators [12]:

$$\tilde{P} = e^{-\Delta}P\,e^\Delta = P + \frac{1}{1!}[P,\Delta] + \frac{1}{2!}[[P,\Delta],\Delta]$$
$$+ \frac{1}{3!}[[[P,\Delta],\Delta],\Delta] + \cdots \quad (8)$$

This equation is valid only if the two matrix exponentials commute, which clearly is true in the natural representation, equation (6). Thus we may write the transformed density matrix as a power series:

$$\tilde{P}^\mu_{\bullet\nu} = e^{-\Delta^\mu_{\bullet\lambda}}P^\lambda_{\bullet\sigma}e^{\Delta^\sigma_{\bullet\nu}} = \sum_{j=0}^\infty \frac{1}{j!}(P^{[j]})^\mu_{\bullet\nu} \quad (9)$$

Here $P^{[0]} = P$, and each higher order contribution follows from the recursion:

$$(P^{[j+1]})^\mu_{\bullet\nu} = [P^{[j]},\Delta]^\mu_{\bullet\nu} = (P^{[j]})^\mu_{\bullet\alpha}\Delta^\alpha_{\bullet\nu} - \Delta^\mu_{\bullet\alpha}(P^{[j]})^\alpha_{\bullet\nu}. \quad (10)$$

To work in the covariant integral representation, where the elements of both $P$ and $\Delta$ are fully contra-

variant, we must change the tensor character of equation (9) term by term:

$$\tilde{P}^{\mu\nu} = e^{-A^{\mu}_{\bullet\alpha}} P^{\alpha\beta} e^{A^{\bullet\nu}_{\beta}} = \sum_{j=0} \frac{1}{j!} (P^{[j]})^{\mu\nu}. \qquad (11)$$

The successive contributions in this equation are defined by transforming equation (10) with metric matrices:

$$(P^{[j+1]})^{\mu\nu} = [P^{[j]}, A]^{\mu\nu} = (P^{[j]})^{\mu\alpha} S_{\alpha\beta} A^{\beta\nu} - A^{\mu\alpha} S_{\alpha\beta} (P^{[j]})^{\beta\nu}. \qquad (12)$$

In this representation, one may further simplify the commutators. Given $X, S$ symmetric and $A$ antisymmetric, we may write

$$[X, A] = XSA - ASX = XSA - (X^{\dagger} S^{\dagger} A^{\dagger})^{\dagger}$$
$$= XSA + (XSA)^{\dagger}. \qquad (13)$$

The entire sequence of commutators is symmetric and fully contravariant, as it must be to update a symmetric contravariant density matrix. The total cost is a single matrix multiply per order in either representation.

### 2.3. *Energy derivative and the steepest descent direction*

Let us consider the gradient of the energy with respect to variations in the density matrix, parametrized according to equation (6) (or equivalently equation (7)). In terms of 1- and 2-electron matrix elements in the natural representation, the SCF energy function for Hartree–Fock theory is (note that trivially one may delete the two-electron terms for tight-binding theory, or modify them with the exchange–correlation potential for Kohn-Sham density functional theory):

$$E = P^{\mu}_{\bullet\nu} H^{\nu}_{\bullet\mu} + \tfrac{1}{2} P^{\mu}_{\bullet\nu} \Pi^{\nu\sigma}_{\bullet\bullet\mu\lambda} P^{\lambda}_{\bullet\sigma}. \qquad (14)$$

The energy derivative of this expression becomes:

$$(\nabla E)^{\nu}_{\bullet\mu} \equiv \frac{\partial E}{\partial A^{\mu}_{\bullet\nu}} = \frac{\partial E}{\partial \tilde{P}^{\lambda}_{\bullet\sigma}} \frac{\partial \tilde{P}^{\lambda}_{\bullet\sigma}}{\partial A^{\mu}_{\bullet\nu}} = F^{\sigma}_{\bullet\lambda} \frac{\partial \tilde{P}^{\lambda}_{\bullet\sigma}}{\partial A^{\mu}_{\bullet\nu}}, \qquad (15)$$

where the SCF Fock matrix $F$ is defined (in the natural representation) as $F^{\nu}_{\bullet\mu} = H^{\nu}_{\bullet\mu} + \Pi^{\nu\sigma}_{\bullet\bullet\mu\lambda} P^{\lambda}_{\bullet\sigma}$. The energy derivatives may be completed (assuming we are initially at $A = 0$) to obtain:

$$(\nabla E)^{\nu}_{\bullet\mu} = F^{\sigma}_{\bullet\lambda} \frac{\partial \tilde{P}^{\lambda}_{\bullet\sigma}}{\partial A^{\mu}_{\bullet\nu}} \bigg|_{A=0} = F^{\nu}_{\bullet\lambda} P^{\lambda}_{\bullet\mu} - P^{\nu}_{\bullet\lambda} F^{\lambda}_{\bullet\mu}. \qquad (16)$$

A derivative in the natural representation, such as equation (16), may be used immediately to define a steepest descent step, because the gradient and the curvy step variables are tensorially consistent in this representation. Thus the steepest descent direction becomes

$$A^{\mu}_{\bullet\nu} = -(\nabla E)^{\mu}_{\bullet\nu}. \qquad (17)$$

On the other hand, if one chooses to express the energy derivative in a different representation, such as the covariant integral representation, the situation is different, as we have emphasized previously [10]. Then the energy derivative is fully covariant:

$$(\nabla E)_{\mu\nu} = \frac{\partial E}{\partial A^{\mu\nu}} = F_{\mu\lambda} P^{\lambda\sigma} S_{\sigma\nu} - S_{\mu\lambda} P^{\lambda\sigma} F_{\sigma\nu}, \qquad (18)$$

which is not tensorially consistent with the contravariant curvy step variables, $A^{\mu\nu}$. Then the correct steepest descent direction (i.e., the vector that is physically the same as defined above in the natural representation) may be obtained only by changing the character of the gradient to fully contravariant using the metric matrices:

$$A^{\mu\nu} = -(\nabla E)^{\mu\nu} = -S^{\mu\lambda} (\nabla E)_{\lambda\sigma} S^{\sigma\nu}. \qquad (19)$$

### 2.4. *Curvy steps*

Having defined a step direction, either through the gradient vector given above, or by some other method, such as conjugate gradients, we must choose the step length $s$ and then evaluate the modified density matrix via the equations of section 2.2. We shall now show that these two tasks can be accomplished together, defining what we call a curvy step. A curvy step will be along a given direction, defined by $A^{\mu}_{\bullet\nu}$, such that the length $s$ of the step minimizes the trial energy associated with a fixed Fock matrix, $E_{\text{trial}} = \tilde{P}^{\mu}_{\bullet\nu} F^{\nu}_{\bullet\mu}$. If this trial energy were fully minimized with respect to all elements of $A^{\mu}_{\bullet\nu}$, this would be equivalent to diagonalizing the matrix, as in a direct diagonalization of a tight-binding Hamiltonian, or a classical Roothaan step in an SCF calculation. However, with our step direction fixed, the most we can do is find the minimum energy along the path so defined.

Employing equation (9) for the transformed density matrix $\tilde{P}$, we may then immediately write a polynomial expansion of $E_{\text{trial}}$ in powers of $s$:

$$E_{\text{trial}} = \sum_{j=0} \frac{s^j}{j!} E^{[j]}_{\text{trial}}, \qquad (20)$$

where the coefficients of this Taylor expansion are simply the traces of the Fock matrix with the density matrix increments defined by equations (9) and (10):

$$E^{[j]}_{\text{trial}} = (P^{[j]})^{\mu}_{\bullet\nu} F^{\nu}_{\bullet\mu}. \qquad (21)$$

The evaluation of equation (21) is at no cost relative to making the density matrix increments, which are required in any case to define $\tilde{P}$. The punchline is that we are able to do a sophisticated (relatively high order polynomial) search along this step direction at no additional cost if the intermediate matrices are saved. If they

are not saved, the cost is still only twice the evaluation of a step of unit length.

The curvy steps have a number of interesting properties. (1) The steps are not straight lines in the space of density matrices, but rather represent curved paths, as a result of sampling only idempotent density matrices with the correct electron number out of the full space of unconstrained density matrices. (2) For a given initial choice of direction, the curvy steps are 'geodesics' in this curved space. In other words they are the proper analogues of straight lines in this curved space in the sense of being a path of minimum distance between two points. This follows directly follows from the results obtained by Edelman and coworkers [13], and requires that each step be defined starting from $\Delta^\mu_{\bullet\nu} = 0$. We have also recently exploited this in a direct minimization algorithm [14]. (3) The curvy steps truncated at $j = 1$ (i.e., terms linear in $\Delta^\mu_{\bullet\nu}$) are straight-line steps in density matrix space, that only preserve idempotency through first order. From an idempotent starting point, this descent direction is exactly the same as in the pioneering Li–Nunes–Vanderbilt (LNV) linear scaling method [15]. At higher orders, the curvy steps are therefore successive generalizations of the LNV scheme that better preserve idempotency. (4) If the curvy step trial energy (and therefore also the density update) is truncated at finite order $M$, then idempotency and electron number will not be retained exactly. These properties can be restored by one or two iterations of the quadratically convergent McWeeny purification transform [16]:

$$\tilde{P}^\mu_{\bullet\nu} = 3P^\mu_{\bullet\lambda}P^\lambda_{\bullet\nu} - 2P^\mu_{\bullet\lambda}P^\lambda_{\bullet\sigma}P^\sigma_{\bullet\nu}. \qquad (22)$$

Geometrically, this corresponds to an orthogonal approach to the idempotent surface [17].

## 3. Toy applications

### 3.1. *Minimal basis HeH$^+$*

We have previously examined HeH$^+$ with the minimal STO-3G basis, and $R = 0.785$ Å as a very simple system for examining SCF (in this case Hartree–Fock) energy minimization in non-orthogonal coordinates [10]. Here we shall use it as a first example to illustrate the nature of curvy step displacements. Using an obvious matrix notation in the covariant integral representation, we take as our initial (idempotent) density matrix $P_0$ the situation with an electron localized on the hydrogen:

$$P_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \qquad (23)$$

Note that the overlap matrix in the same basis is not diagonal, but has an off-diagonal element. The off-diagonal element of the $2 \times 2$, anti-symmetric $\Delta$ matrix is the only free parameter describing changes in the density matrix:

$$\Delta = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix}. \qquad (24)$$

Figure 1 plots the energy as a function of the off-diagonal element for increasing expansion orders (1, 2, 3, ..., 25). Steps using the seventh and higher order



Figure 1. Energy of the minimal basis HeH$^+$ model problem as a function of the off-diagonal element of $\Delta$ for increasing expansion orders (1, 2, 3, ..., 25). Steps using the seventh and higher order expansions remain uniformly idempotent from the starting $P_0$ to the global minimum of the system ($x = -1.1538$). Higher order expansions show a periodic structure in the energy functional.

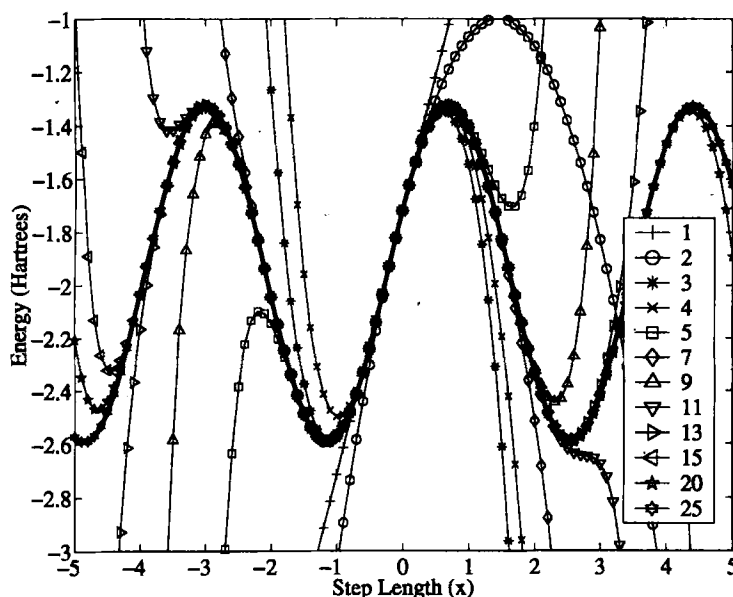expansions remain uniformly idempotent from starting at $P_0$ to the global minimum of the system ($x = -1.1538$). The energy plots for the higher order expansions show the beginnings of a periodic structure. This structure is consistent with the view of equation (1) producing 'rotations' within the basis that preserve the scalar product. For an infinite order expansion, each of the periodic energy minima in $x$ represents an identical density matrix.

Figure 2 shows the diagonal $P_{11}$ matrix element along the $x$ axis, and the off-diagonal $P_{12}$ density matrix element along the $y$ axis as functions of the off-diagonal element of $\Delta$. Each frame of this figure

shows the curvy step expressions truncated to a different expansion order (i.e., truncating equation (11) at order $j$, for various values of $j$). The ellipse within each frame represents all the possible diagonal and off-diagonal density matrix elements that correspond to an idempotent density matrix. Thus all physical density matrices must fall on this ellipse, and by inspecting the deviation of the trial density matrices along a curvy step, we can get a feel for the extent to which idempotency is preserved along a curvy step of order $j$. While stopping at $j = 1$ gives linear paths (like the LNV method), the higher order truncations are nonlinear curves. With increasing order, the curvature
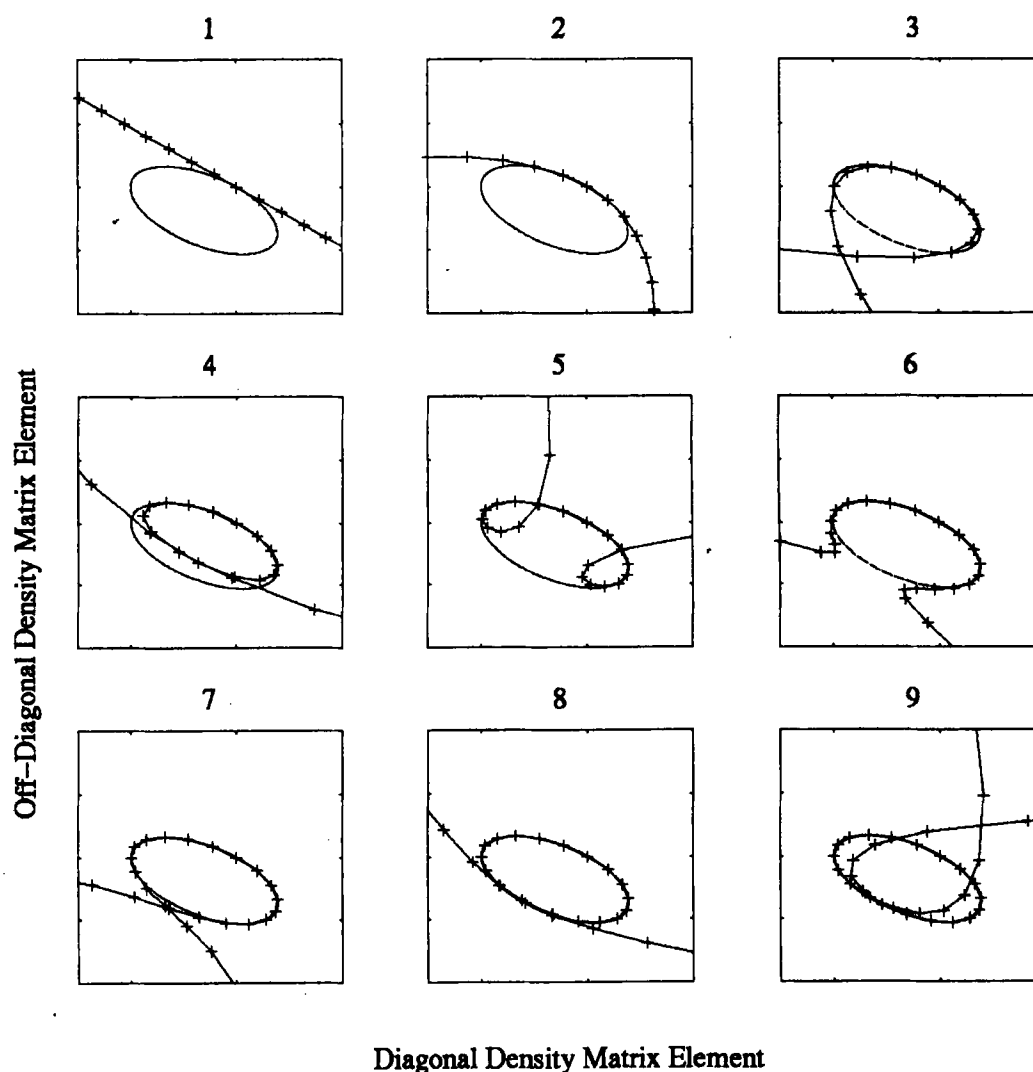


**Diagonal Density Matrix Element**

Figure 2. Diagonal density matrix element ($P_{11}$, for the H site), and the off-diagonal density matrix element of the density matrix for the steps displayed in figure 1. Each inset shows the step for a curvy step truncated at a different expansion order. The ellipse within each frame represents diagonal and off-diagonal density matrix elements for an idempotent density matrix. Linear changes in $\Delta$ appear as curved paths in this density matrix representation. Within increasing order, the curvy steps more closely track the idempotent ellipse. The periodicity in the energy shown in figure 1 corresponds to multiple trips around the ellipse in this figure.

more closely tracks the idempotent ellipse. The periodicity of the energy described in figure 1 corresponds to multiple trips around the ellipse, which is beginning to emerge by the 9th order.

We can make several general observations from this simple model system. The use of high order curvy steps permits one to take longer steps without such strong loss of idempotency as occurs at lower order (in particular at linear order). For relatively low orders of expansion, the energy minima obtained by searching along a finite order curvy step direction do not correspond exactly to the minimum on the exactly idempotent surface. The minimum obtained for a low order curvy step is a nearly idempotent point closer to the final $\bar{P}$. It may be purified, which on the figures shown corresponds to an orthogonal approach [17] to the ellipse of idempotency.

## 3.2. *Purifying curvy steps*

Repeated applications of a $j$th order curvy step will rapidly produce non-idempotent density matrices. For this reason, if the result from a curvy step derived from a low order expansion is used as a new starting $P$, the purification transform must first be applied to ensure an idempotent $P$. In this section we explore the ability to purify a non-idempotent $P$ that arises from taking curvy steps of various orders. We take a model system described by a $50 \times 50$ tight-binding type Hamiltonian with $-2$ self interactions and $-1$ nearest neighbour coupling as appropriate for a linear chain. The overlap between nearest neighbours was 0.1. Thus,

$$H_{ij} = -2\delta_{ij} - \delta_{(i+1)j} - \delta_{(i-1)j}, \qquad (25)$$

$$S_{ij} = \delta_{ij} + 0.1\delta_{(i+1)j} + 0.1\delta_{(i-1)j}. \qquad (26)$$
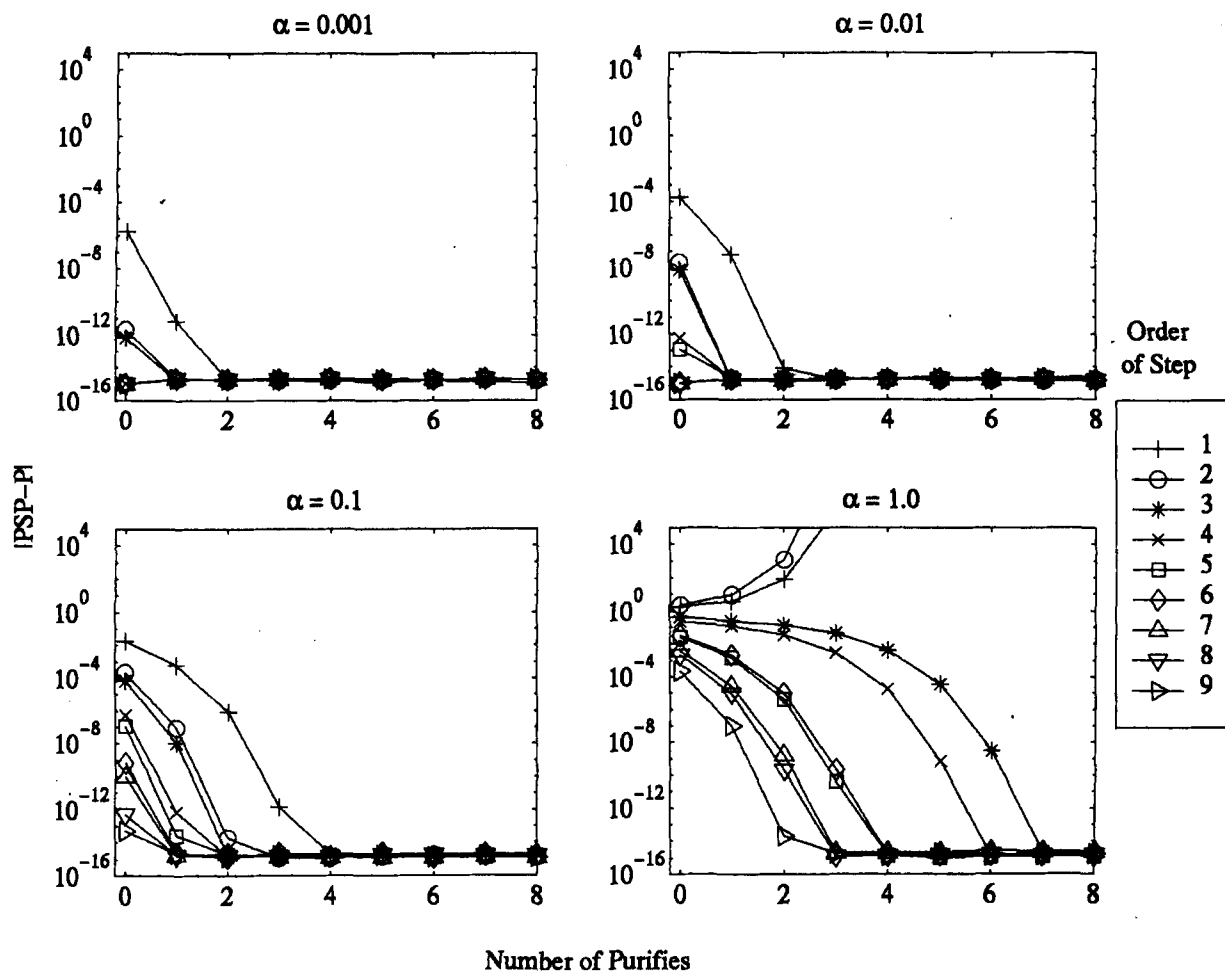


Figure 3. Deviation from idempotency (measured as the maximum value of |PSP − P|) as a function of the curvy step expansion order for the tight-binding model problem. The data were obtained by taking a step along the negative gradient (premultiplied by $\alpha$) from an idempotent starting point using various expansion orders. Each frame represents a different length step with $\alpha = 1.0$. representing a full step along the negative gradient. A first-order step corresponds to an LNV steepest descent step. The idempotent starting point is far from the true minimum.

The initial idempotent density matrix was constructed using a recurrence based density matrix construction method that Schmidt-orthogonalizes local vectors on each site. The data were collected by taking a curvy step (of various orders) along the negative gradient of the energy, and then purifying the resulting $\hat{P}$. The length of the step is determined by the energy gradient pre-multiplier $\alpha$. The purification becomes more difficult as one increases the length of the step ($\alpha$) or decreases the order of the curvy step. The $\alpha = 1$ step simulates the initial step in a standard steepest descent procedure.

For low expansion orders ($j = 1$ and $j = 2$) and $\alpha = 1$, the curvy step proceeds outside the convergence range of the purification transform, equation (22). In these cases, an idempotent density matrix cannot be recovered as seen from the divergence of the deviation from idempotency (figure 3). Any minimization method taking such a step would fail to converge. However, when using the high order expansions (or small step sizes), the steps quickly purify to idempotent points. For the very high order expansions, essentially no purification is necessary. By contrast, the linear step associated with the LNV method would correspond to a first-order expansion on this graph, and thus would fail to converge for a full steepest descent step.

This suggests again that the use of curvy steps in non-orthogonal coordinate systems will allow one to take longer steps than is possible in an LNV type method. The gradient itself remains identical to the LNV gradient (at least at an idempotent point). However, instead of stepping linearly along the gradient, a curvy step is taken. This could potentially reduce the total number of steps to the minimum, as well as providing stronger protection against 'runaway solutions', which correspond to moving outside the range of the purification transform.

## 4. Conclusion

We have presented a tensor formulation for unitary transformations of the density matrix in basis sets that may be either orthogonal, non-orthogonal, or biorthogonal. The tensor approach makes the connection between different representations very clear, and generalizes the recent atomic orbital basis presentation of [5]. We have demonstrated the connection between this method and the widely used Li–Nunes–Vanderbilt (LNV) linear scaling method. Steps in the LNV method correspond to the truncating the general expressions given here at first order in the expansion variables. Thus LNV steps are straight-line steps, while higher order truncations yield curved steps through the space of density matrices, which are successively more accurate approximations to the true geodesic in the search direction. The nature of the curvy steps obtained by truncating at higher than first order is illustrated by calculations on two toy model problems, which employ tight-binding Hamiltonians. It appears that longer displacements are possible than with straight line steps. We are currently exploring the use of curvy steps to develop new linear scaling methods that exploit this potential.

## References

[1] ROOTHAAN, C. C. J., 1951, *Rev. mod. Phys.*, **23**, 69.
[2] WHITE, C. A., JOHNSON, B. G., GILL, P. M. W., and HEAD-GORDON, M., 1996, *Chem. Phys. Lett.*, **253**, 268.
[3] SCUSERIA, G. E., 1999, *J. phys. Chem.* A, **103**, 4782.
[4] GOEDECKER, S., 1999, *Rev. mod. Phys.*, **71**, 1085.
[5] HELGAKER, T., LARSEN, H., OLSEN, J., and JORGENSEN, P., 2000, *Chem. Phys. Lett.*, **327**, 397.
[6] BACKSAY, G. B., 1981, *Chem. Phys.*, **61**, 385.
[7] HUTTER, J., PARRINELLO, M., and VOGEL, S., 1994, *J. chem. Phys.*, **101**, 3862.
[8] MASLEN, P. E., OCHSENFELD, C., WHITE, C. A., LEE, M. S., and HEAD-GORDON, M., 1998, *J. phys. Chem.* A, **102**, 2215.
[9] ARTACHO, E., and DELBOSCH, L. M., 1991, *Phys. Rev.* A, **43**, 5770.
[10] WHITE, C. A., MASLEN, P., LEE, M. S., and HEAD-GORDON, M., 1997, *Chem. Phys. Lett.*, **276**, 133.
[11] HEAD-GORDON, M., MASLEN, P. E., and WHITE, C. A., 1998, *J. chem. Phys.*, **108**, 616.
[12] WEISS, G. H., and MARADUDIN, A. A., 1962, *J. math. Phys.*, **3**, 771.
[13] EDELMAN, A., ARIAS, T. A., and SMITH, S. T., 1998, *SIAM J. Matrix Anal. Appl.*, **20**, 303.
[14] VAN VOORHIS, T., and HEAD-GORDON, M., 2002, *Molec. Phys.*, **100**, 1713.
[15] LI, X. P., NUNES, W., and VANDERBILT, D., 1993, *Phys. Rev.* B, **47**, 10891.
[16] MCWEENY, R., 1960, *Rev. mod. Phys.*, **32**, 335.
[17] BOWLER, D. R., and GILLAN, M. J., 1999, *Comput. Phys. Commun.*, **120**, 95.