

New techniques of self-consistent field theory

Trygve Helgaker

Simen Reine, Erik Tellgren

Centre for Theoretical and Computational Chemistry, University of Oslo, Norway

Stinne Høst, Thomas Kjærgaard, Branislav Jansík, Poul Jørgensen, Jeppe Olsen

Lundbeck Foundation Center for Theoretical Chemistry, University of Aarhus, Denmark

TACC 2008

Theory and Applications of Computational Chemistry 2008

Songjiang New Century Grand Hotel Shanghai, China

September 22–27, 2008

## Newton's method and self-consistent field (SCF) theory

- The calculation of electronic energies is a classical problem of optimization:

“Before 1940 relatively little was known about methods for numerical optimization of functions in many variables. There had been some least squares calculations carried out, and steepest descent type methods had been applied in some physics problems. The **Newton method** in many variables was known, and more sophisticated methods were being attempted such as the **self-consistent field** for variational problems in theoretical chemistry.”

R. Fletcher, *Practical Methods of Optimization*, Wiley, Chichester, 1980.

- We shall here explore **the relationship between Newton's method and the SCF method**
  - Hartree–Fock (HF) and Kohn–Sham (KS) energies
- The traditional SCF optimization method consists of **two alternating steps**:
  - **the Roothaan–Hall (RH) diagonalization step**
  - **Pulay's DIIS density-averaging step**
  - both steps are needed for convergence, which is nevertheless sometimes difficult
- We shall see how it is possible to merge the two steps into **a single concerted step**
  - **the augmented Roothaan–Hall (ARH) step**
  - more robust convergence

## Hartree–Fock and Kohn–Sham energies

- We are here interested in **energy optimization** in HF and KS theories

$$E(\mathbf{D}) = \text{Tr } \mathbf{D}\mathbf{h} + \frac{1}{2} \text{Tr } \mathbf{D}\mathbf{G}(\mathbf{D}) + E_{\text{XC}}(\mathbf{D})$$

- we use standard notation, with the two-electron part given by

$$G_{pq}(\mathbf{D}) = \sum_{rs} \langle \chi_p \chi_q | r_{12}^{-1} | \chi_r \chi_s \rangle D_{rs} + \text{exchange}$$

- we will ignore the exchange–correlation term
- we assume an orthonormal basis (Cholesky or Löwdin AOs) but not MOs

- For an  $N$ -electron system, there are certain constraints on the **density matrix**

$$\mathbf{D}^T = \mathbf{D}, \quad \text{Tr } \mathbf{D} = N, \quad \mathbf{D}^2 = \mathbf{D}$$

- parameterization of the density matrix in terms of orbitals:

$$\mathbf{D} = \mathbf{C}\mathbf{C}^T, \quad \mathbf{C}^T\mathbf{C} = \mathbf{I}$$

- parameterization of the density matrix in terms of an antisymmetric matrix:

$$\mathbf{D}(\mathbf{X}) = \exp(-\mathbf{X})\mathbf{D}_{\text{ref}}\exp(\mathbf{X}), \quad \mathbf{X}^T = -\mathbf{X}$$

- We begin by reviewing the standard SCF optimization scheme

## The Roothaan–Hall (RH) diagonalization step

- In the  $n$ th iteration, we introduce the **Fock matrix** constructed from  $\mathbf{D}_n$ :

$$\mathbf{F}_n = \mathbf{h} + \mathbf{G}(\mathbf{D}_n), \quad \delta\mathbf{D} = \mathbf{D} - \mathbf{D}_n$$

- In terms of this Fock matrix, we may rewrite the energy as:

$$E(\mathbf{D}) = \underbrace{\text{Tr } \mathbf{D}\mathbf{h}}_{\text{1-electron}} + \frac{1}{2} \underbrace{\text{Tr } \mathbf{D}\mathbf{G}(\mathbf{D})}_{\text{2-electron}} = \underbrace{\text{Tr } \mathbf{D}\mathbf{F}_n}_{\text{large}} + \frac{1}{2} \underbrace{\text{Tr } \delta\mathbf{D}\mathbf{G}(\delta\mathbf{D})}_{\text{small}} - \underbrace{\frac{1}{2} \text{Tr } \mathbf{D}_n \mathbf{G}(\mathbf{D}_n)}_{\text{constant } k_n}$$

- Assuming that  $\delta\mathbf{D}$  is small, we neglect the term quadratic in  $\delta\mathbf{D}$ :

$$E(\mathbf{D}) \approx E_{\text{RH}}(\mathbf{D}) = \text{Tr } \mathbf{D}\mathbf{F}_n + k_n \leftarrow \text{Roothaan–Hall energy}$$

- To minimize the RH energy, we use the parameterization  $\mathbf{D} = \mathbf{C}\mathbf{C}^T$  and obtain

$$E_{\text{RH}}(\mathbf{D}) = \text{Tr } \mathbf{C}^T \mathbf{F}_n \mathbf{C} + k_n$$

- variation in  $\mathbf{C}$  under the constraints  $\mathbf{C}^T \mathbf{C} = \mathbf{I}$  yields the RH eigenvalue problem:

$$\mathbf{F}_n \mathbf{C} = \mathbf{C}\boldsymbol{\epsilon}$$

- In the RH step, we perform an **exact minimization of an approximate HF energy**
  - it may or may not converge upon iteration

### Pulay's DIIS step

- To obtain the  $n$ th DIIS step, we begin with the same energy decomposition:

$$E(\mathbf{D}) = \underbrace{\text{Tr } \mathbf{D} \mathbf{F}_n}_{\text{large}} + \frac{1}{2} \underbrace{\text{Tr } \delta \mathbf{D} \mathbf{G}(\delta \mathbf{D})}_{\text{small}} + k_n$$

- However, we now **retain the quadratic term** but **linearize the density matrix**:

$$\delta \mathbf{D} = \exp(-\mathbf{X}) \mathbf{D}_n \exp(\mathbf{X}) - \mathbf{D}_n \approx [\mathbf{D}_n, \mathbf{X}], \quad \mathbf{X}^T = -\mathbf{X}$$

- variation with respect to  $\mathbf{X}$  now gives the stationary condition:

$$[\mathbf{D}_n, \mathbf{F}_n + \mathbf{G}(\delta \mathbf{D})] = \mathbf{0}$$

- To avoid the high cost of the  $\mathbf{G}$  term, we expand  $\delta \mathbf{D}$  in old density matrices:

$$\delta \mathbf{D} \approx \sum_i c_i (\mathbf{D}_i - \mathbf{D}_n) \quad \Rightarrow \quad \mathbf{G}(\delta \mathbf{D}_n) = \sum_i c_i (\mathbf{F}_i - \mathbf{F}_n)$$

- the stationary condition now becomes

$$\sum_{i=0}^n [\mathbf{D}_n, c_i \mathbf{F}_i] = \mathbf{0}, \quad \sum_{i=0}^n c_i = 1 \quad \leftarrow \text{approx. satisfied by norm min.}$$

- In the DIIS step, we perform a **quasi-Newton step with a linearized density matrix**
  - it corrects the RH step for the neglect of the  $\mathbf{G}$  term

## The traditional RH–DIIS method

- The traditional SCF optimization **alternates** between two steps

$$E(\mathbf{D}) = \underbrace{\text{Tr } \mathbf{D} \mathbf{F}_n}_{\text{large}} + \frac{1}{2} \underbrace{\text{Tr } \delta \mathbf{D} \mathbf{G}(\delta \mathbf{D})}_{\text{small}} + k_n$$

- **RH step:** minimize  $E(\mathbf{D}(\mathbf{X}))$  by linearizing the energy dependence on  $\delta \mathbf{D}$
- **DIIS step:** minimize  $E(\mathbf{D}(\mathbf{X}))$  by linearizing the density-matrix dependence on  $\mathbf{X}$
- linearization means neglect of second-order (Hessian) information
- In none of these steps do we make a full use of all available Hessian information
  - the iterations may therefore oscillate, diverge or converge to an incorrect solution
  - we shall now see how it is possible to combine these separate steps into one
- To achieve this, we return to the full energy expression as a function of  $\mathbf{X}$ 
  - we shall straightforwardly apply Newton’s method
- Learning from SCF theory, we then simplify the Newton step
  - a single step that merges the RH and DIIS steps into one
  - the **augmented Roothaan–Hall (ARH) method**

## Newton's method applied to the SCF energy

- We expand the full SCF energy in orders of  $\mathbf{X}$  around  $\mathbf{D}_n$ :

$$\mathcal{E}(\mathbf{X}) = \text{Tr } \delta \mathbf{D} \mathbf{F}_n + \frac{1}{2} \text{Tr } \delta \mathbf{D} \mathbf{G}(\delta \mathbf{D}) + k_n$$

$$\delta \mathbf{D} = [\mathbf{D}_n, \mathbf{X}] + \frac{1}{2} [[\mathbf{D}_n, \mathbf{X}], \mathbf{X}] + \dots, \quad \mathbf{X}^T = -\mathbf{X}$$

- We truncate the expansion at **second order** and minimize within the **trust region**

$$\|\mathbf{X}\| \leq h_n \quad \leftarrow \text{trust-region radius}$$

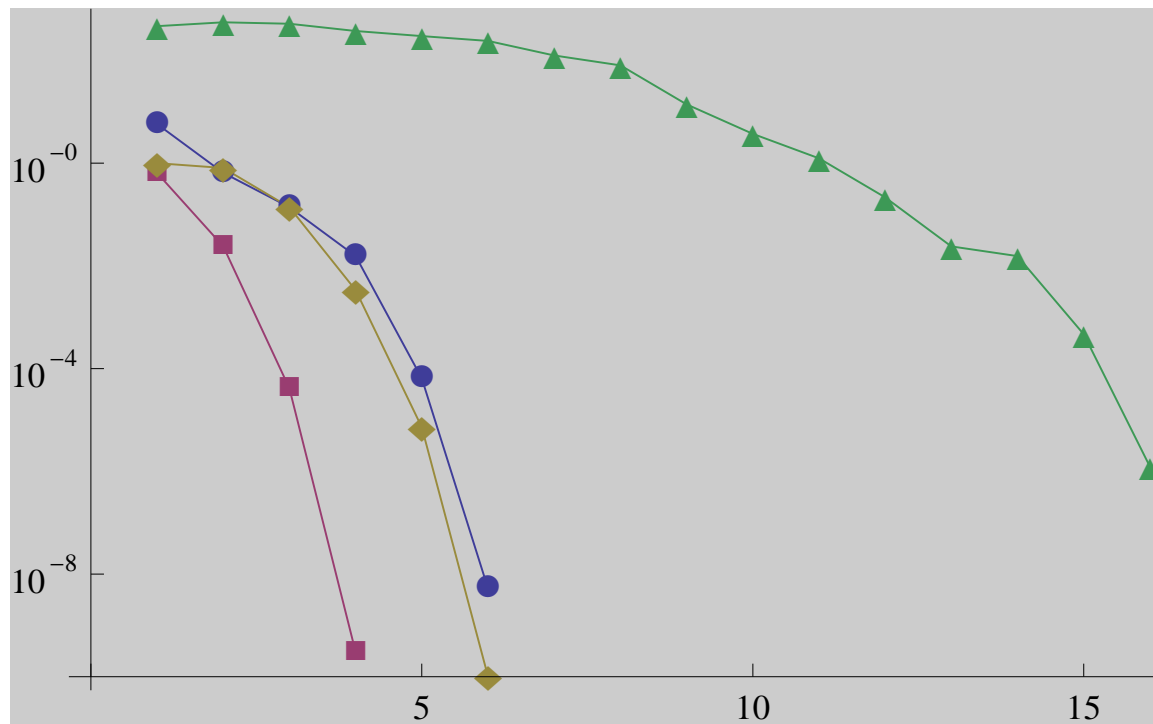
- This constrained minimization leads to the level-shifted Newton step

$$\sum_{\nu} \left( \frac{\partial^2 \mathcal{E}(\mathbf{X})}{\partial X_{\mu} \partial X_{\nu}} - \lambda \delta_{\mu\nu} \right) X_{\nu} = - \frac{\partial \mathcal{E}(\mathbf{X})}{\partial X_{\mu}}$$

- **gradient** and **Hessian** carry information about **slope** and **curvature**, respectively
- the **level-shift parameter**  $\lambda$  is adjusted to restrict the search inside the trust region
- Newton's method has been applied to the SCF energy by other authors
  - G. B. Bacskay (1981), H. J. Aa. Jensen and H. Ågren (1984)
  - we modify the density matrix  $\mathbf{D}(\mathbf{X})$  directly
  - no MOs are introduced and diagonalization is avoided

## Some examples of Newton optimizations

- Newton density-matrix optimizations of the total electronic energy
  - water HF/DZ (red), water dimer B3LYP/DZ (yellow), adenine HF/STO-3G (blue), rhodium complex HF/STO-3G+ADZ (green)



- Convergence fast in the local region, slow but smooth in global region
  - global convergence guaranteed by the trust-region method
  - a large number of Fock/KS matrix reevaluations are needed: 24, 42, 54, 211



## The structure of the Newton matrix equation

- The structure of the Newton matrix equation reflects that of the energy:

$$\underbrace{(\mathbf{F}_n^{\text{vv}} - \mathbf{F}_n^{\text{oo}}) \mathbf{X} + \mathbf{X} (\mathbf{F}_n^{\text{vv}} - \mathbf{F}_n^{\text{oo}})}_{\text{large 2nd-order F term}} + \underbrace{\mathbf{G}^{\text{ov}}([\mathbf{D}_n, \mathbf{X}]) - \mathbf{G}^{\text{vo}}([\mathbf{D}_n, \mathbf{X}])}_{\text{small 2nd-order G term}} = \underbrace{\mathbf{F}_n^{\text{vo}} - \mathbf{F}_n^{\text{ov}}}_{\text{1st-order F term}}$$

- projections onto the occupied and virtual orbital spaces

$$\mathbf{F}_n = \mathbf{F}_n^{\text{oo}} + \mathbf{F}_n^{\text{ov}} + \mathbf{F}_n^{\text{vo}} + \mathbf{F}_n^{\text{vv}}, \quad (\mathbf{P}_o = \mathbf{D}_n, \mathbf{P}_v = \mathbf{I} - \mathbf{D}_n)$$

- note: in the MO basis, the first term consists of orbital-energy differences

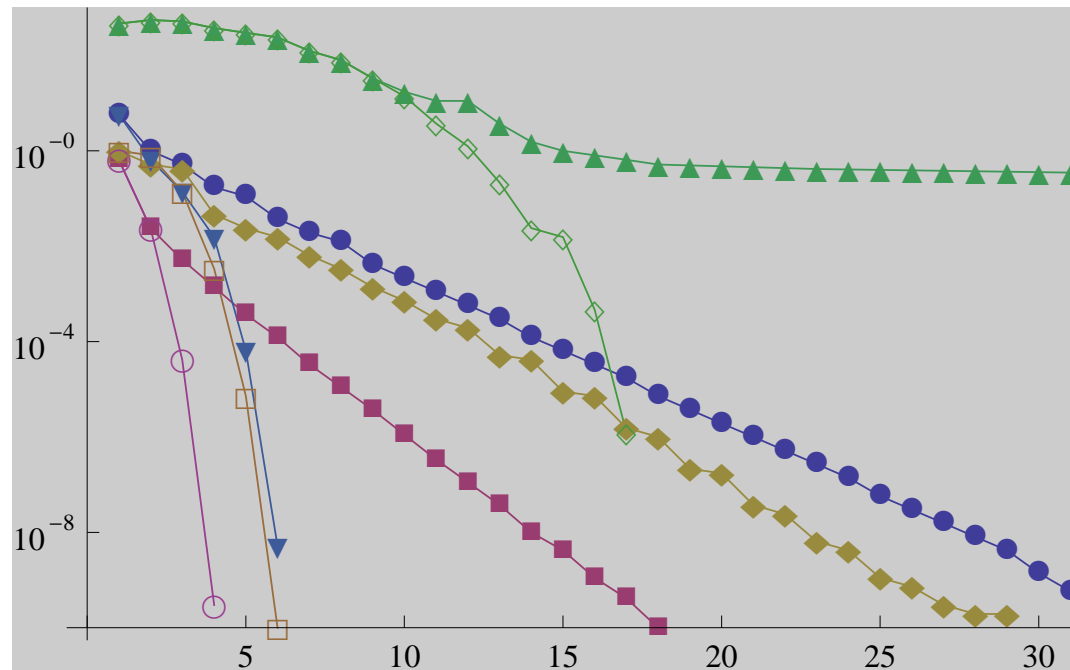
- The Newton matrix equation is solved **iteratively**, in about ten iterations
  - at each iteration, we evaluate the left-hand side with a trial  $\mathbf{X}$  matrix
- The (typically large) **F contribution** to the Hessian transformation is **inexpensive**
  - requires only matrix multiplications
- The (typically small) **G contribution** to the Hessian transformation is **expensive**
  - requires a new two-electron integral evaluation
- Perhaps we can **approximate** the G term, in the spirit of DIIS?
  - let us first drop the G term altogether, as done in the RH step!

## A (very) simplified Newton method

- Neglecting the G term, we obtain the following simple matrix equation

$$(\mathbf{F}_n^{\text{vv}} - \mathbf{F}_n^{\text{oo}}) \mathbf{X} + \mathbf{X} (\mathbf{F}_n^{\text{vv}} - \mathbf{F}_n^{\text{oo}}) = \mathbf{F}_n^{\text{vo}} - \mathbf{F}_n^{\text{ov}}$$

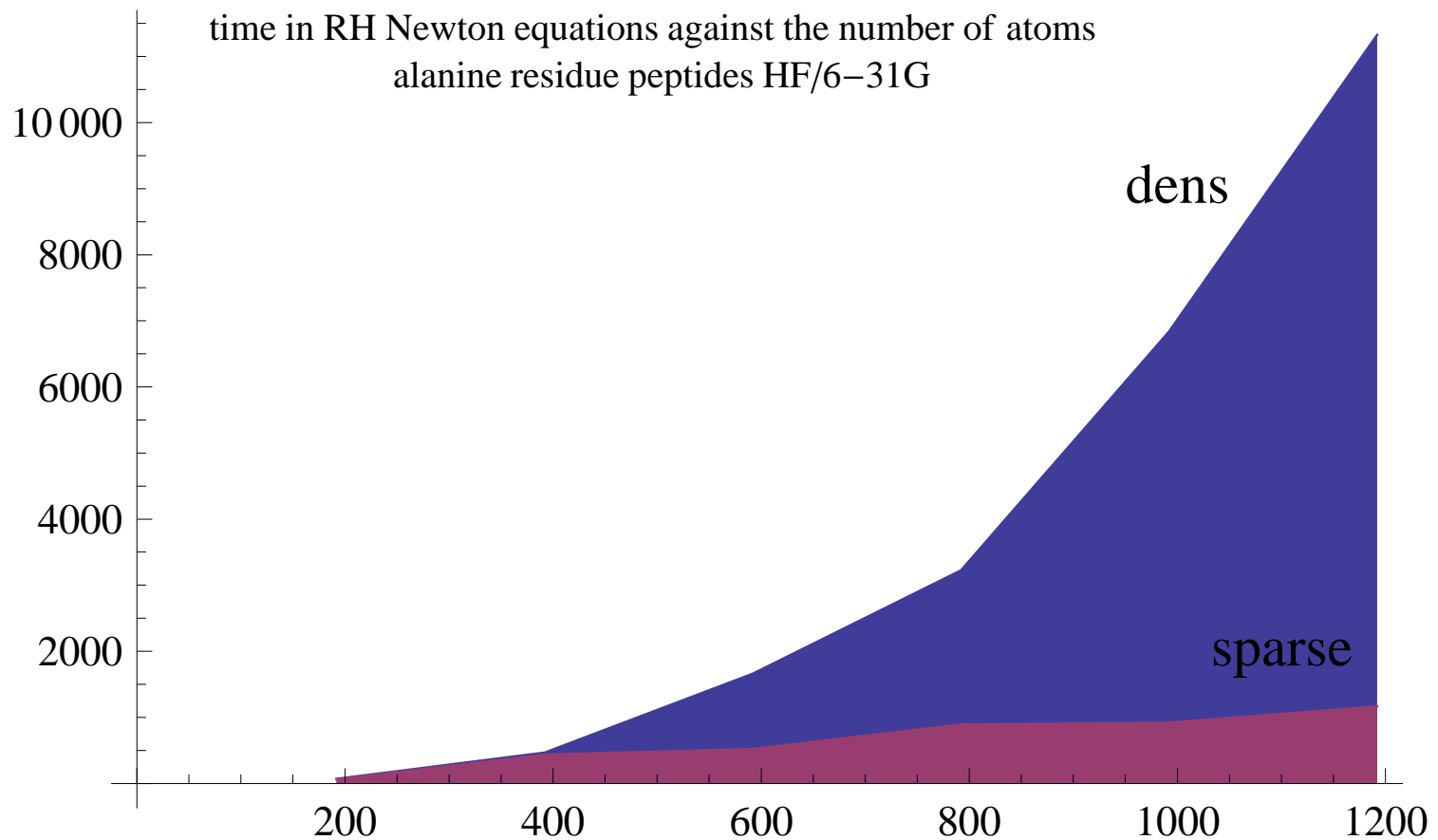
- But energy convergence suffers since the Newton step is now approximate
  - water (red), water dimer (yellow), adenine (blue), rhodium complex (green)



- **linear convergence** in the local region, global convergence not guaranteed
- the number of Fock/KS matrix reevaluations: 20 (24), 50 (42), 33 (54),  $\infty$  (211)

## The Roothaan–Hall Newton method

- A **byproduct**: by removing the G term, we are minimizing the RH energy
  - we are in fact carrying out a **Roothaan–Hall diagonalization** step!
  - an alternative to diagonalization for large systems
  - 50–100 sparse matrix multiplications



## The augmented Roothaan–Hall method

- Consider the expensive part of the SCF electronic Hessian

$$\mathbf{H}_n(\mathbf{X}) = \mathbf{G}^{\text{ov}}([\mathbf{D}_n, \mathbf{X}]) - \mathbf{G}^{\text{vo}}([\mathbf{D}_n, \mathbf{X}])$$

- During the optimization, we have collected a set of density matrices

$$\mathbf{D}_{in} = \mathbf{D}_i - \mathbf{D}_n$$

- in these directions, we invoke the **quasi-Newton assumption** (exact in HF theory)

$$\mathbf{G}(\mathbf{D}_{in}) = \mathbf{G}(\mathbf{D}_i) - \mathbf{G}(\mathbf{D}_n) \approx \mathbf{F}_i - \mathbf{F}_n = \mathbf{F}_{in} \leftarrow \text{available}$$

- To evaluate the Hessian contribution approximately, we introduce a projector onto  $\mathbf{D}_{in}$ :

$$\mathbf{P}_n(\mathbf{D}_{in}) = \mathbf{D}_{in}, \quad \mathbf{P}_n^2 = \mathbf{P}_n$$

- we now obtain the following approximate Hessian contribution (JO):

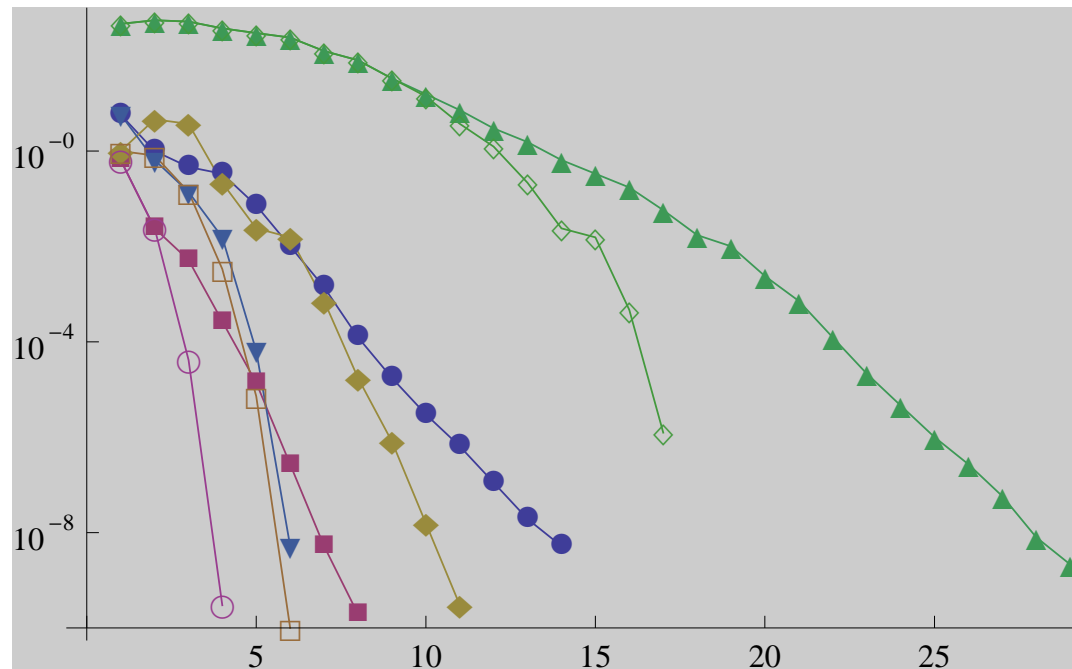
$$\bar{\mathbf{H}}_n(\mathbf{X}) = \sum_{ij} (\mathbf{F}_{in}^{\text{ov}} - \mathbf{F}_{in}^{\text{vo}}) [\mathbf{T}^{-1}]_{ij} \text{Tr}(\mathbf{D}_{jn}[\mathbf{D}_n, \mathbf{X}]), \quad T_{ij} = \text{Tr} \mathbf{D}_{in} \mathbf{D}_{jn}$$

- no new Fock/KS matrix evaluations are needed

- We term the resulting method the **augmented Roothaan–Hall (ARH) method**
  - a quasi-Newton method with the dominant part of the Hessian treated exactly

## Examples of ARH optimizations

- The ARH method is similar to Newton globally but slower locally
  - water (red), water dimer (yellow), adenine (blue), rhodium complex (green)



- It requires the **fewest number of Fock/KS evaluations** (for the four systems):

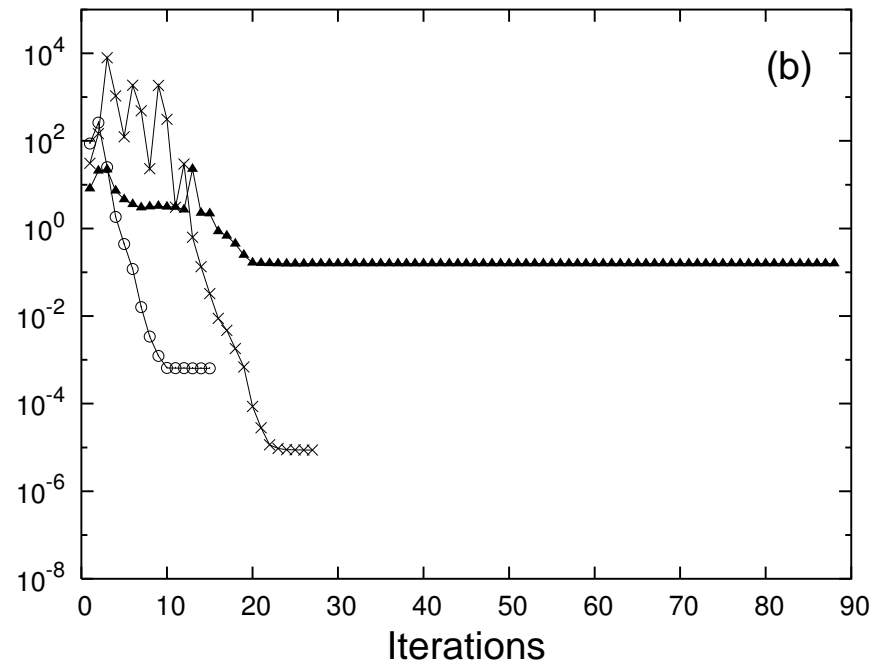
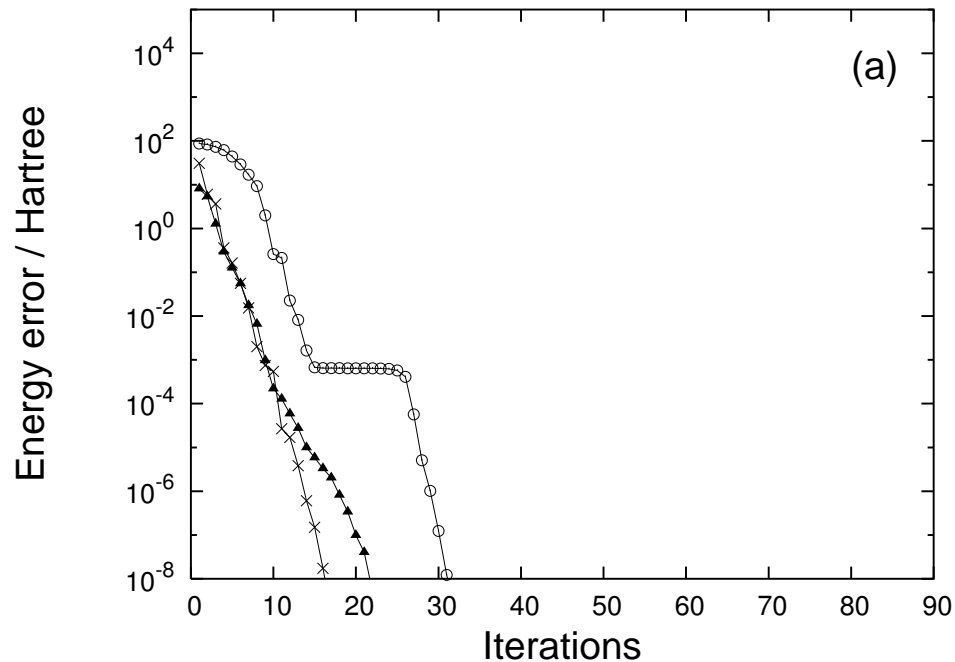
Newton	24	42	54	211
RH	20	50	33	$\infty$
ARH	9	15	15	30

## The performance of the ARH algorithm

- The ARH method appears to be **robust** for optimizations
  - it treats the dominant part of the Hessian exactly (like the RH step)
  - it treats the remainder in a quasi-Newton fashion (like the DIIS step)
  - it makes **full use of all available Hessian information in a single, concerted step**
- In the following, we shall compare it with the RH–DIIS method
  - convergence to gradient norm smaller than  $10^{-4}$
  - HCORE of Hückel starting guess
- Høst *et al.*, PCCP 2008, 10, 5344–5348

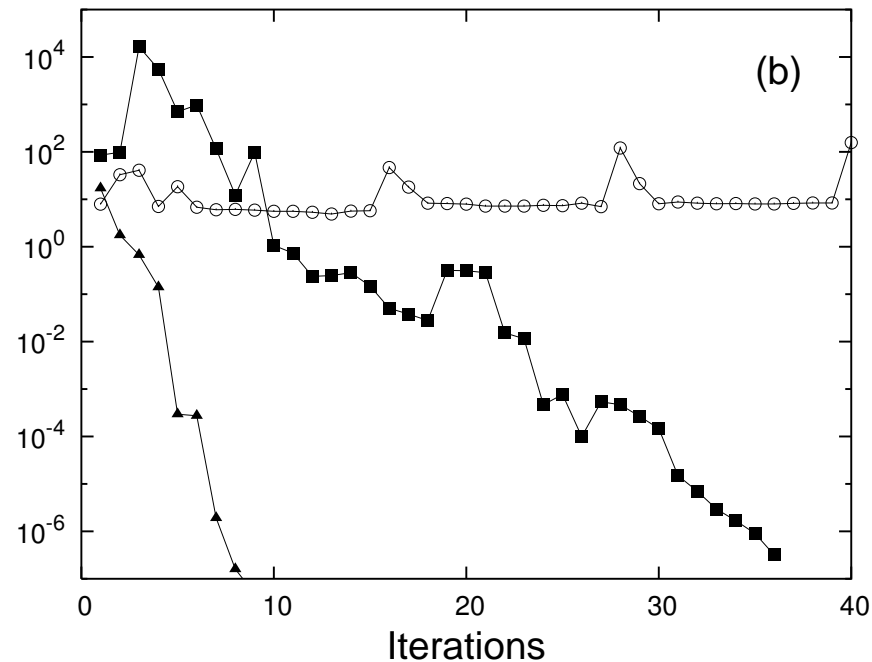
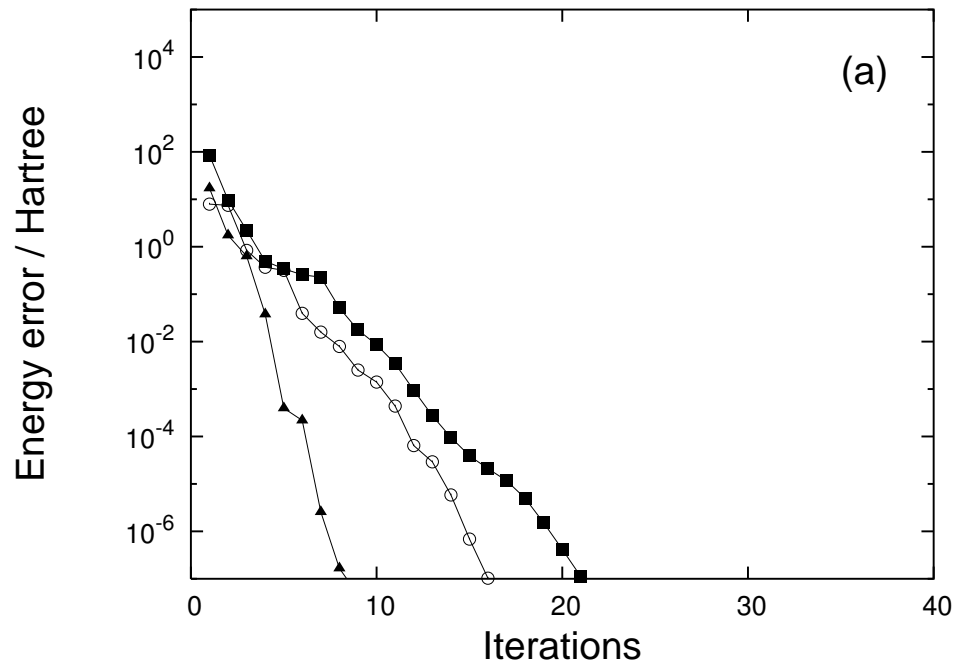
## The ARH method compared with RH-DIIS (I)

- Convergence for a variety of molecules
  - cadmium–imidazole complex (circles) B3LYP/3-21G (10/89)
  - 29-residue polyalanine peptide (crosses) B3LYP/6-31G (292/1599)
  - model vitamin B12 (triangles) HF/AhlichVDZ (74/428)
- Energy-error on a logarithmic scale (ARH left, RH-DIIS right)
  - RH-DIIS converges to a saddle point, ARH always to a minimum
  - standard programs perform in the same manner



## The ARH method compared with RH-DIIS (II)

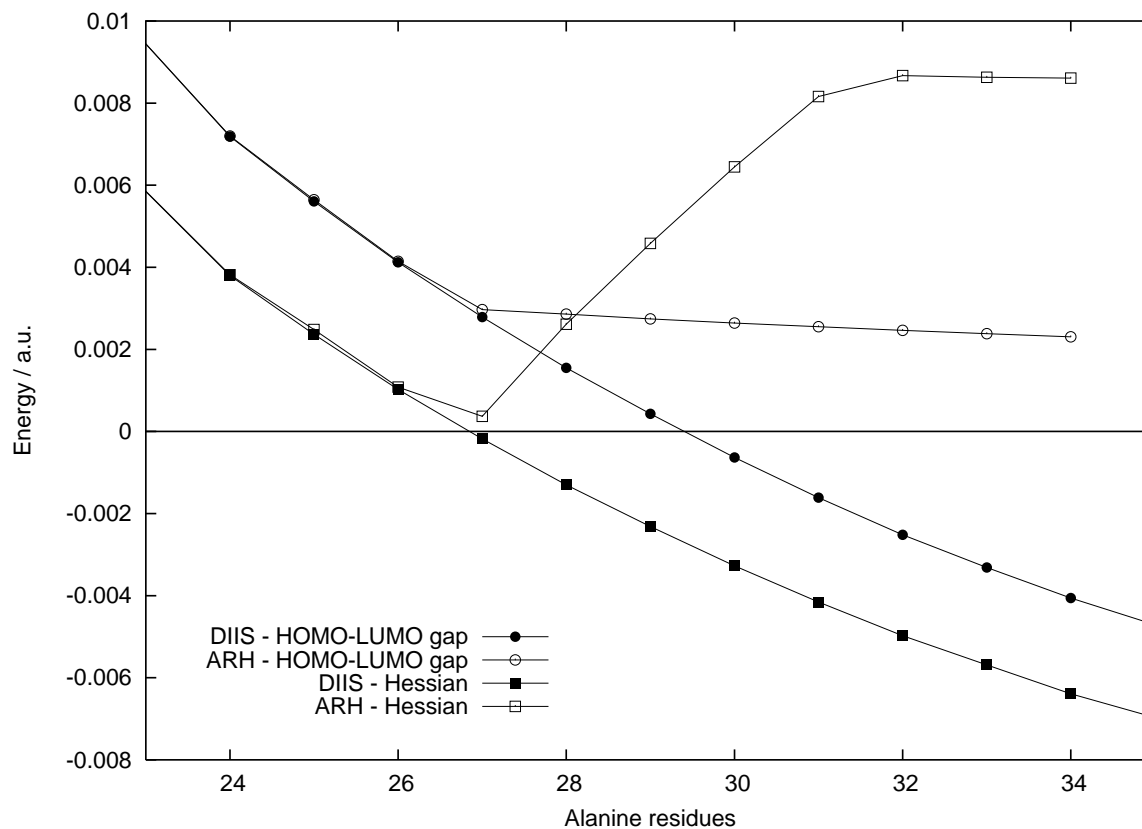
- Convergence of calculations for a variety of molecules
  - cluster of 51 water molecules (full triangles) B3LYP/cc-pVTZ (153/2958)
  - insulin (full squares) B3LYP/6-21G (786/4417)
  - model vitamin B12 (empty circles) BP86/AhrlrichsVDZ (74/428)
- Energy-error on a logarithmic scale (ARH left, RH-DIIS right)
  - ARH converges smoothly in all cases
  - RH-DIIS good for water cluster, oscillates for insulin, diverges for vitamin B12



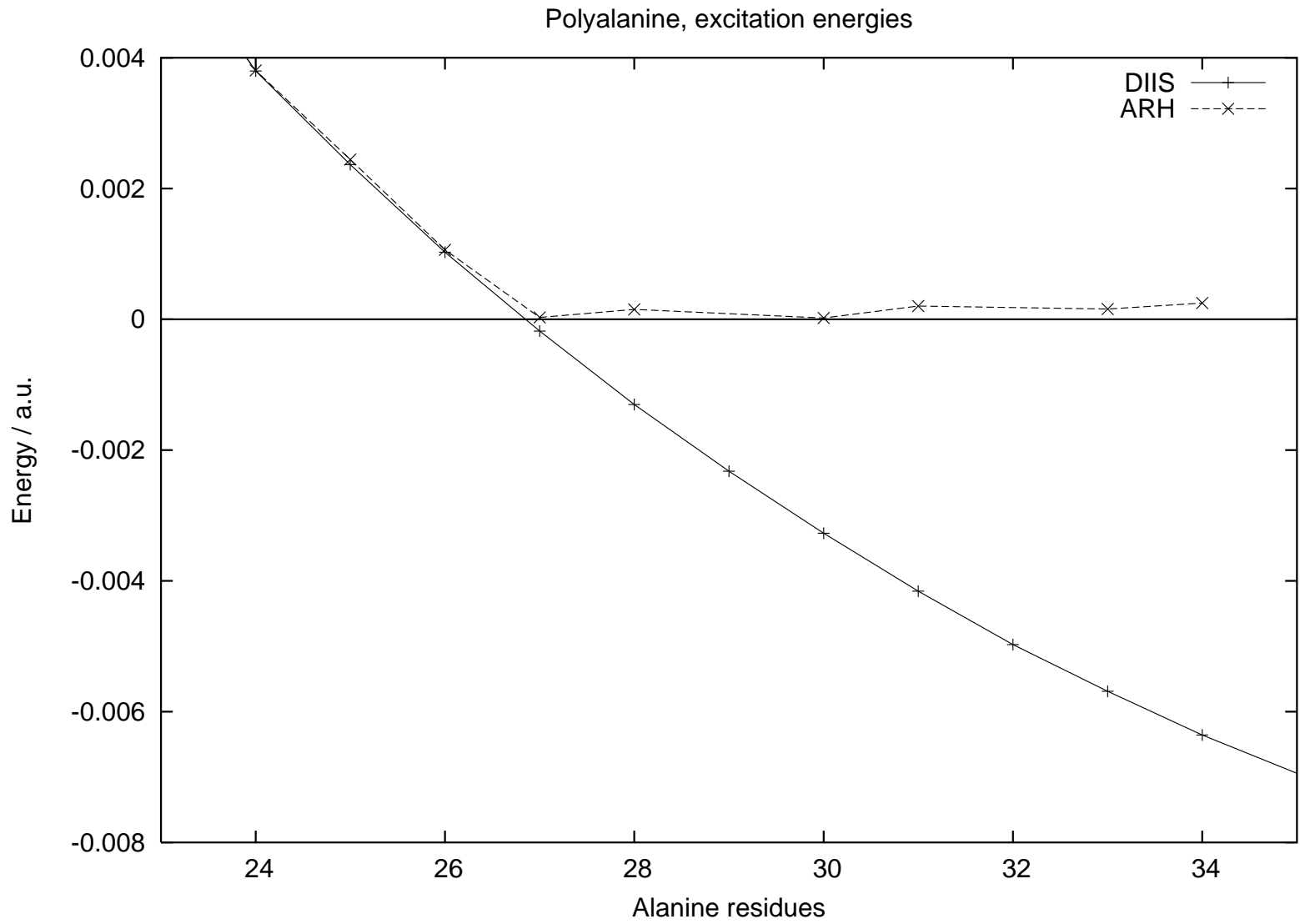


## Optimization of polyalanine peptides (B3LYP)

- The RH-DIIS method converges well for less than 29 alanine residues
  - for larger peptides, a level shift of 0.01 enforces convergence
  - however, the solution then becomes a saddle point
  - the same behavior observed for standard programs
- The ARH method converges well for all systems



B3LYP/6-31G excitation energies of alanine residue peptides



### Three-level optimization

- The ARH scheme has recently been implemented within a **three-level (3L) scheme**:
  - grand-canonical atomic optimization
  - valence-basis molecular optimization
  - full-basis molecular optimization
- Comparison with other codes for 23 transition-metal complexes
  - test set by van Lenthe *et al.* (2006), for their ATOMS scheme

	number of iterations		
	best (%)	aver. (%)	worst (%)
van Lenthe <i>et al.</i>	100	100	100
<b>ARH-3L</b>	<b>64</b>	<b>76</b>	<b>100</b>
QCP1	91	109	141
QCP2	114	157	205

- B. Jansík *et al.*, submitted

## Conclusions

- The traditional SCF scheme alternates between two steps
  - the Roothaan–Hall diagonalization step
  - the DIIS averaging step
  - neither step makes full use of the available Hessian information
- In the augmented RH (ARH) method, the two steps are merged into one concerted step
  - the dominant part of the Hessian (orbital-energy differences) is treated exactly
  - the remainder is treated accurately in the space of the previous density matrices
  - trust-region method used for step-size control
  - no diagonalization necessary—sparsity may be fully exploited

## ICQC 2009 satellite symposium

- Molecular properties: bridging the gap between theory and experiment
  - Oslo, Norway, June 18–21 2009