

**ReadMe file for new version of mixer (VER10.8)**

**L. D. Marks, 29 January 2021**

## **Contents**

### **MIXER (adding and mixing of charge densities)**

Suggested documentation (for latest version)

### **Introduction to Version 10.8 (Peach)**

### **Introduction to Version 9.4 (Mango)**

### **Introduction to the Multisecant Mixers**

### **Parallel atomic minimization algorithms (a.k.a. the Energizer Bunny)**

Brief description of new algorithms that “mix” atomic positions and densities

### **Regularization**

Algorithm for regularization the Jacobian

### **Trust Region Control**

Step limitation

### **Charge Renormalization in MSR1a**

### **What are the Forces in MSR1a?**

### **Ghost Bands and Ghost Forces**

Potential problems with iterative modes

### **Dealing with Hard Problems**

Some explanation and suggestions

### **Undocumented options**

Some controls for the expert or code developer

### **Cases where MSR1a works well**

### **Older Version Notes:**

### **Major Changes with Version 4.1**

### **Major Changes with Version 5.3**

### **Changes with Version 6.0**

### **Changes with Version 7.1.2**

### **Changes with Version 8.0**

## MIXER (adding and mixing of charge densities)

In `mixer` the electron densities of core, semi-core, and valence states are added to yield the total new (output) density (in some calculations only one or two types will exist). Proper normalization of the densities is checked and enforced. As it is well known, simply taking the new densities leads to instabilities in the iterative SCF process. Therefore it is necessary to stabilize the SCF cycle. In *WIEN2k* this is done by mixing the output density with the (old) input density to obtain the new density to be used in the next iteration. Several mixing schemes are implemented, but we mention only:

1. Straight mixing as originally proposed by Pratt (52) with a mixing greed  $Q$

$$\rho_{\text{new}}(r) = (1 - Q)\rho_{\text{old}}(r) + Q\rho_{\text{output}}(r)$$

2. A Multi-Secant mixing scheme contributed by L. Marks (see Marks and Luke 2008), in which all the expansion coefficients of the density from several preceding iterations (usually 6-10) are utilized to calculate an optimal direction in each iteration. This version is by far superior to the other schemes making them quite obsolete. It is robust and stable (works nicely also for magnetic systems with 3d or 4f states at EF) and usually converges at least 30 % faster than the old BROYD scheme.
3. Two new variants on the Multi-Secant method including a rank-one update (see Marks 2013) which appear to be faster and equally robust

At the outset of a new calculation (for any changed computational parameter such as k-mesh, matrix size, lattice constant etc.), any existing `case.broydx` files must be deleted (since the iterative history which they contain refers to a ``different`` incompatible calculation).

If the file `case.c1msum_old` cannot be found by `mixer`, a ``PRATT-mixing`` with mixing greed 1.0 is done.

*Note: a `case.c1mval` file must always be present, since the LM values and the K-vectors are read from this file.*

The total energy and the atomic forces are computed in `mixer` by reading the `case.scf` file and adding the various contributions computed in preceding steps of the last iteration. Therefore `case.scf` must not contain a certain ``iteration-number``

more than once and the number of iterations in the scf file must not be greater than 999.

For LDA+U calculations `case.dmatup/dn` and for hybrid-DFT (switch -eece) `case.vorbup/dn` files will be included in the mixing procedure. With the new mode MSR1a (or MSECa) atomic positions will also be mixed (effectively optimized).

## 1 Execution

The program `mixer` is executed by invoking the command:

```
mixer mixer.def Of x mixer [-eece]
```

A spin-polarized case will be detected automatically by `x` due to the presence of a `case.clmvalup` file. For an example see `fccNi` (sec. [10.2](#)) in the *WIEN2k* package.

## 2 Dimensioning parameters

The following parameters are collected in file `param.inc`, :

NCOM	number of LM terms in density
NRAD	number of radial mesh points
NSYM	order of point group
Traptouch	minimum acceptable distance between atoms in full optimization model

## 3 Input

Below a sample input (written automatically by `1start`) is provided for  $\text{TiO}_2$  (rutile), one of the test cases provided with the *WIEN2k* package.

```
----- top of file: case.inm -----
MSR1  0.d0  YES  (PRATT/MSEC1 background charge (+1 for additional e), NORM
      0.2      MIXING GREED
1.0 1.0      Not used, retained for compatibility only
999 8        nbroyd nuse
----- bottom of file -----
```

Interpretive comments on this file are as follows:

**line 1:**

(A5,\*)

switch, bgch, norm

- switch MSR1 Recommended: A Rank-One Multisecant that is slightly faster than MSEC3 in most cases. For MSR1a see later.
- MSEC3 Multi-Secant scheme (Marks and Luke 2008). This is equivalent to DIIS, with trust regions added.
- MSEC4 Similar to MSEC3 (above), but mixes the higher LM values inside spheres by an adaptive PRATT scheme. This leads to a significant reduction of program size and file size (`case.broyd*`) for unit cells with many atoms and low symmetry (factor 10-50) with only slightly worse mixing performance.
- MSR2 A variant of MSR1 which only mixes the L=0 values inside the spheres, similar to MSEC4
- PRATT Pratt's scheme with a fixed greed
- PRAT0 Pratt's scheme with a greed restrained by previous improvement, similar to MSEC3
- bgch Background charge for charged cells (+1 for additional electron, -1 for core hole, if not neutralized by additional valence electron)
- norm YES Charge densities are normalized to sum of Z
- NO Charge densities are not normalized

**line 2:**

free format

- greed mixing greed Q. Essential for PRAT0 and PRATT, rather less important for Multisecant methods. Q is automatically reduced by the program. In case that the scf cycle fails due to large charge fluctuations, this can be further reduced but this can lead to stagnation. One should rarely reduce this below 0.05.

### line 3 (optional):

(free format)

f\_pw, f\_clm

f\_pw        Not used, retained for compatibility only.

f\_clm        Not used, retained for compatibility only..

### line 4 (optional):

(free format)

nbroyd, nuse

nbroyd        Not used, retained for compatibility only.

nuse        For all the Multisecant Methods: Only nuse prior steps are used (this value has some influence on the optimal convergence. Usually 6-10 seems reasonable and 8 is recommended). For MSR1a of large cells sometimes 16 is better.

### Line 5 or more (optional), additional switches:

#### *For Hard Problems:*

**STIFF**        **Recommended** if you run into severe convergence problems, this will probably solve them

**STIFFER**      A version of **STIFF** that goes a little further

**TRAD X**        Reduces the maximum atom movement to X au from the default 0.1. For hard problems **TRAD 2E-2** may be useful

**BLIM X**        Reduces the upper step bound **X**, for instance to **1.0**

**FAST**         May be faster for well-conditioned problems

#### *Other:*

**PRATT**        Uses **PRATT** mode with the step size control of **MSR1**

**LAMBDA X**     Fixes the regularization value to **X**, for instance 1D-4

**VERBOSE**      Outputs additional information that may not be useful for most people

**LESS**         Reduces the amount of information **VERBOSE** produces

**RESTART X**    Restarts from a converged density with an initial **GREED** of **X**

**ELIN**         Mixing of global linearization energy, can be useful for d or f electrons

**DIIS**         The DIIS (PULAY, ANDERSON) method, with are unscaled MSEC3. Not recommended.

**MSGB**         Multisecant Good Broyden, not recommended.

**MSEC3**        The same as MSEC3 on the first line. (DIIS/PULAR/ANDERSON/MSGB also)

## Control files

.msec	The existence of the file .msec forces the next step to use a mixing greed which is read from the file. The code deletes the file if it is present. In some cases the algorithm might get trapped with small greed terms (e.g. 0.01), and this allows it to be reset higher so avoiding Wolfe condition starvation.
.pratt	The existence of the file .pratt forces the next step to be a Pratt step, without a restart. If the file contains a number, this is the mixing greed used. The code deletes the file if it is present.
.Blimit	Increases the trust region to value in the file for the next iteration – do grep :TRUST case.scf
.Climit	Increases the charge trust region of single atoms to the value in the file for the next iteration – do grep :TRUST case.scf
.CTOlimit	Increases the charge trust region of all atoms to the value in the file for the next iteration – do grep :TRUST case.scf
.PWlimit	Increases the plane wave trust region to the value in the file for the next iteration – do grep :TRUST case.scf
.ATlimit	Increases the charge trust region for atoms movement to the value in the file for the next iteration – do grep :TRUST case.scf
.Fast	Presence uses faster (less safe) controls for the next step
.Restart	Restarts from the prior density, preserving the trust region radii and exploiting prior information about the GREED.

## Introduction to Version 10.8 (Peach)

Version 10.8 has important changes:

1. The previous variable scaling of the atom positions, orbital potentials/density matrix and gradients has been removed. This has the disadvantage that the scaling in absolute terms may be a little inconsistent. However, this is more than outweighed by the advantage of having a constant merit function.
2. All trust regions are now completely controlled internally using a combination of improvement and (more importantly) estimations based upon the previous step. This is much more robust. This is also done for the GREED.
3. MSR1 now uses a more aggressive choice of Slambda. In addition, the regularization has been simplified. Both make the prediction algorithms more robust.
4. If the algorithm makes a very bad step, it backtracks and attempts to correct itself with a Pratt-like step and throws away the last step. This is to avoid contamination by bad steps.
5. If the algorithm makes a somewhat bad step, it still uses this step but backtracks so it does a better job.
6. Addition of a new **STIFF** mode (and **STIFFER**), which has a high probability (probably more than 90%) of converging in all cases so long as a disastrous ghost-band problem is not encountered.
7. Addition of DIIS/Pulay/Anderson and Good-Broyden methods (slower or less stable).
8. It has experimental constraints added, which are described in the separate files **README\_Constraints**. These are still somewhat experimental, so should be used with care. In some cases they work well. In others, particularly searches for transition states they do not always work as well as desired – yet.
9. In the very first iteration, a PRATT step, the pseudocharge within each sphere is backprojected onto a L=M=0 radial density, and a correction is then applied. What this does is to significantly reduce the pseudocharge which can significantly improve the convergence for problems involving d electrons and many other cases where the initial atomic densities are poor. It will also improve the convergence when changing RKMAX, since there is a significant coupling between RKMAX and the pseudocharge.
10. The old MSEC1 has been retired.
11. Many code simplifications and condensations have been made. The code now uses openmp to improve the speed.
12. Adjusted heuristics for what occurs when the user inputs a GREED < 0.2.

## Introduction to Version 9.4 “Mango”

The Mango release represents a major change in some of the inner workings of the mixer, which from all tests to date lead to a factor of 2-3 improvement in the speed for certain cases, and much higher stability. Hopefully this will be written up for a literature publication in the near future, these are just a few preliminary (and incomplete) notes.

For routine calculations, most users will see little difference in the behavior of this version and the one currently in the WIEN2k\_17 distribution (version 8.0), and the MSEC1 from the previous version is preserved intact. In some cases it will be significantly faster, and also much more stable. The GREED used can be very different from what was in previous versions – and should be much better in general. In some cases it will be smaller, even down to 0.0001, in some it will be larger. The scaling of the unpredicted step (Beta) is now calculated much more accurately, which can lead in some cases to much smaller (but better) steps.

The largest change is that the L2 normalization of all the variables in the previous versions has been replaced. The mixing of the plane waves as well as the density within the muffin tins is now done in units of  $\text{Density} \times \text{Volume}^{1/2}$ . These units automatically scale all the variables so that they have approximately equal importance, which makes the Jacobian closer to a diagonal matrix. With these units the density within the muffin tins is now divided by the radial distance, which will improve convergence of densities closer to the nucleus. This scaling is not perfect, and for different cases the “ideal” scaling may differ from it by a factor of 2-3, but it is close.

The second, very significant change is in the estimation of the mixing greed and the scaling beta of the predicted step. In earlier versions implicit methods were used to estimate these using the relative improvement. While this was adequate, it was not optimal.

With this release the algorithm now calculates both of these explicitly using stored information from the previous steps. The mixing GREED that the algorithm now uses can be very, very different from what it was before, ranging from 10-100 times smaller to significantly larger. All indicators are that these estimates are rather good, better than what most users (or even experts) can do. I suspect that I probably could not pick better values for the GREED and beta in most cases.

This release also includes significant improvements to the Trust Region control. In addition to controls based upon the change in positions and charge change within any muffin tin, it also controls the total change in the plane waves and in all the muffin tins. These increase significantly the stability. Another innovation, is that the algorithm now uses indicators of “trouble” to contract the Trust Regions. At present this is implemented for significant changes in the potential (via the :VZER\* terms in case.sco0) and also QTL-B warnings.

## Major Changes with Version 9.4 “Mango”

This is an incomplete list of the changes

1. No independent scaling of the density by conversion to units of  $\rho dV^{1/2}$ , with
  - a. Renormalization of the CLMs by  $(dx/4\pi R)^{1/2}$
  - b. Renormalization of the PWs by  $V^{1/2}$
2. Storage of the prior unpredicted step in case.broyd1.
3. Estimation of the GREED by calculating what should have been the value for the last step, and using this to help update the value for the current step.
4. Estimation of the Beta parameter (scaling of the predicted step), by estimating what the value should have been for the last step, and using this to help update the value for the current step.
5. Calculation of the diagonal elements of the Jacobian as an update, stored in brody1. These are now used to determine the scaling of for atoms and orbital terms (instead of the old L2 norm).
6. Separate the termination criterion in MSR1a into one on forces, one on movement and change slightly the user output.
7. Addition of a trust region for the largest change in PW components for large spacings.
8. Automatic contraction of the trust region when the code is doing well, to reduce problems when electronic phase changes occur.
9. Addition of a trust region for the total change in the CLM components.
10. Code to contract the trust regions if the potential is changing too much or if there is a QTL-B warning. This may be expanded.
11. New .restart option retaining prior trust radii and mixing information. Can be used when changing from MSR1 to MSR1a (or the other way). Clean operation when changing mode, so rm case.bro\* may be obsolete.
12. Added back full Jacobian scaling (AUTO) and residue scaling (OLD).
13. Cured bugs with gfortran for orbital potentials, and removed creation of unwanted case.dummy files. Cured scaling bug for -eece with versions 16.1 and 17.1
14. The option -DDEBUG turns on a lot of additional information that is probably only of use to LDM
15. Updated some of the documentation

## Algorithm Notes

The algorithm used can be approximately described as the following steps

1. Read old and new density for current step, normalizing and organizing.
2. Check for trouble indicators such as QTL-B warnings or large potential changes. If needed reduce the trust region radii.
3. Pack density into vectors, including orbital terms and atomic forces/positions if needed.
4. Retrieve the history of previous vectors (value and residue).
5. Retrieve history of various parameters such as improvement, whether the step was limited, trust region radii.
6. Scale orbital terms if needed.
7. Calculate some norms.
8. Estimate the value of Lambda in the MSR1 algorithm (if used).
9. Estimate new parameters what the code should have used for the last step
10. Calculate the full Newton step.
11. Check if this step lies within the trust region. If not, contract until it does.
12. Store information used in the algorithm.
13. Unpack the densities and other variables and print them out.

### The trust radii are:

1. That the total step is not too large
2. That the change in any of the smaller (in reciprocal space) plane wave components is not too large
3. That the total change of the charge within any RMT is not too large
4. That the total change of the charge in the RMT summed over all atoms is not too large
5. That the change in atomic positions is not too large
6. That the atoms do not come too close

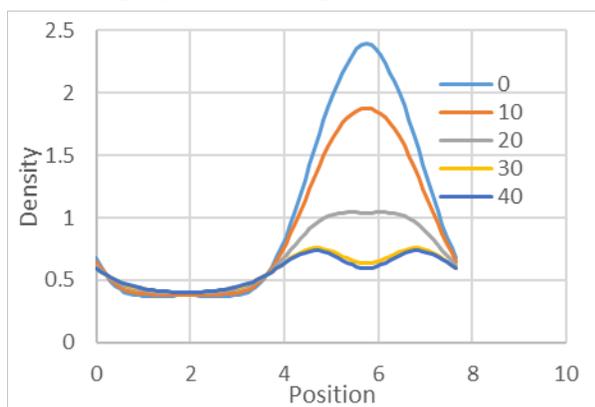
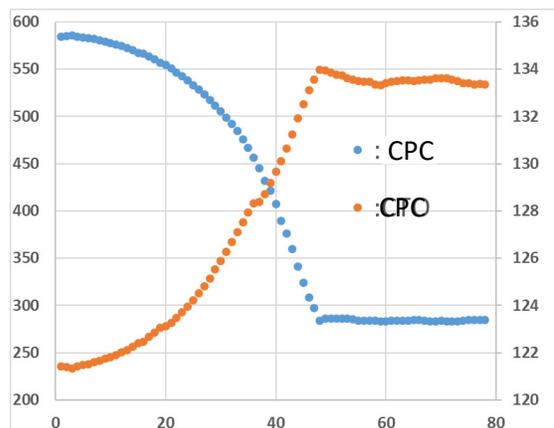
The values of the trust radii can be shown by doing "grep case.scf :TRUST". All are updated within the code based upon predictions for the last step.

## Pseudo Charge

A feature of the Wien2k code is that the plane waves within the RMTs are not directly relevant to any properties, although they play a role in the numerical derivative for the exchange-correlation potential. These densities are part of the total variable set used by the mixer and are important for both the operation of MSR1a and the mixing of the densities.

Unfortunately the initial estimates for these from **dstart** are often not that good, particularly for heavier elements. While the default values for the occupancies of different orbitals in **Istart** are reasonable, often they are too localized and may have the wrong spin state. One can improve convergence, often very significantly, by carefully adjusting the values in **case.inst**. For instance, when working on a problem involving surfaces or large supercell first run a small calculation for the bulk unit cell, then look at the occupancies and spin state and use these. This does require significant experience with using Wien2k, so is not recommended for the novice.

The convergence initially can be slow as the irrelevant plane wave density slowly converges; this may occur even with well-adjusted initial values. Starting from release **Mango** the total Pseudo Charge can be monitored by doing a grep on **:CPC**, **:OPC** or **:NPC** which are, respectively, the total Pseudo Charge at the end of the scf iteration, that of the prior iteration and that predicted from lapw2. Once the Pseudo Charge is converged the algorithm can work better, but there is often a discontinuity. An example of the convergence with iterations of **:CPC** (blue, left axis) and **:CTO** (orange, right axis) is shown in the Figure on the right for an eighty four atom gold surface.

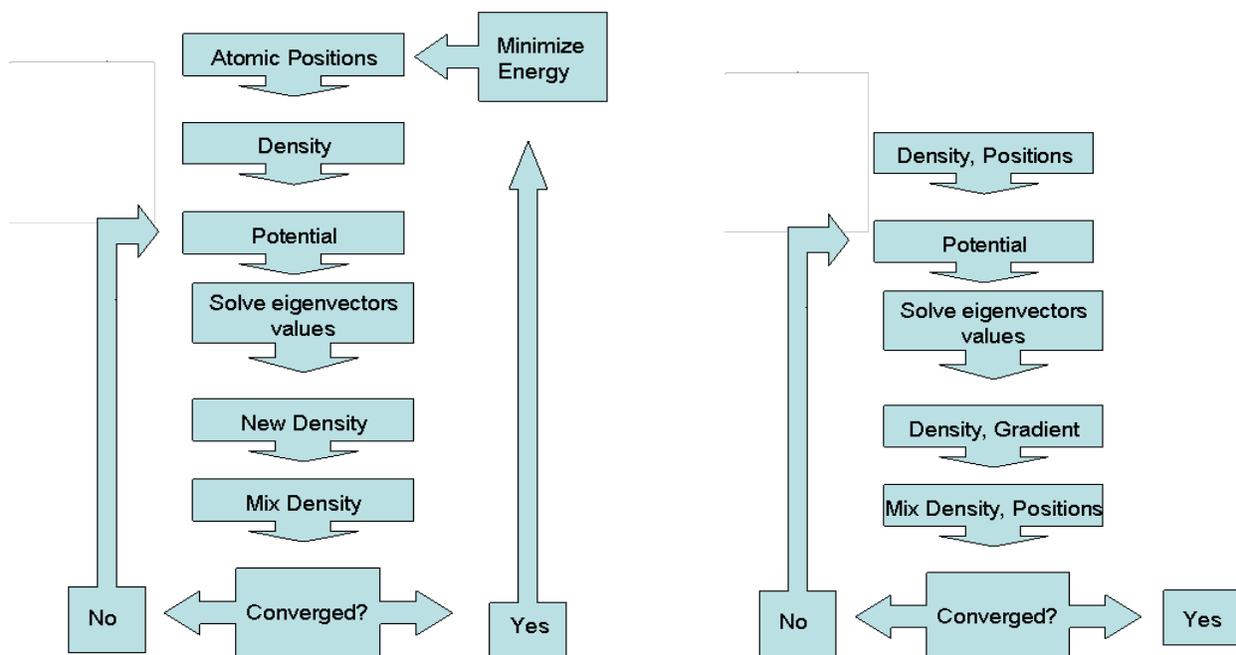


The plane wave density inside the spheres starts with an inappropriate distribution, and only slowly adjusts. The plane wave only density at a W atom is shown for a Fe/W superstructure on the left, for several of the iterations starting from the dstart output. The initial “s-like” density takes ~30 iterations to adjust.

The large error in the Pseudo Charge can drive too large changes in the relevant density, leading to QTL-B warnings or ghostbands at the beginning of the iterations. In addition, the phase transition when the Pseudo Charge converges also has the potential to lead to QTL-B warnings of ghostbands as the Jacobian has changed. Note that with 10.4 there is an improvement in the first (Pratt) iteration which overcomes many of these problems.

## Introduction to the Multisecant Mixers and MSR1a (Updated 2020)

A major innovation with this series of releases is a new algorithm MSR1 (MultiSecant Rank One) as well as an algorithm for simultaneous “mixing” of densities and atomic positions, MSR1a. In a standard DFT calculation, a double loop is used for finding the lowest energy structure as shown below on the left. What MSR1a does is treat the densities and atom positions as variables, and find the self-consistent solution for the densities and gradients (i.e. gradients small or zero) simultaneously, as illustrated below on the right.



This is faster by a factor of 1.5 to 4.0 than the double loop in large test cases, perhaps even more in general.

This algorithm used is not very different from the older MSEC1 or the new MSEC3. Using standard notation, we define two vectors  $s_n$  and  $y_n$  for the  $n$ 'th iteration as

$$s_{n-1} = \rho_n - \rho_{n-1} \quad ; \quad y_{n-1} = R_n - R_{n-1} \quad (1)$$

where the residual is  $R_n = F(\rho_n) - \rho_n$ , i.e. the difference between the density at the start of iteration  $n$  ( $\rho_n$ ) and that ( $F(\rho_n)$ ) produced by solving the Kohn-Sham equations for the potential generated from  $\rho_n$ . We further define two matrices containing some number ( $m$ ) of prior steps:

$$S = [s_{n-m}, s_{n-m+1}, \dots, s_n] \quad \text{and} \quad Y = [y_{n-m}, y_{n-m+1}, \dots, y_n] \quad (2)$$

We describe the original version (MSEC1 as well as MSEC3) via a Jacobian  $H$ :

$$H = \beta (I - Y(Y^T A + \lambda I)^{-1} A^T) + S(Y^T A + \lambda I)^{-1} A^T, \quad \text{where} \quad (3)$$

$\beta$  = mixing greed,  $\lambda$  = regularization term  
 $A = Y$  for the MSEC1 family

The next step is then

$$\rho_{n+1} = \rho_n + H R_n \quad (4)$$

For the MSR1 family, equation (3) is changed to:

$$H = \beta (I - Y \{\text{Reg}(Y^T A)\} A^T) + S \{\text{Reg}(Y^T A)\} A^T \quad (5)$$

Where  $A = (\alpha S + Y)$ , and  $\text{Reg}(Y^T A)$  is a Tikonov regularization of the  $m \times m$  matrix  $Y^T A$ , which can be written as

$$\text{Reg}(Y^T A) = \text{Reg}(B) = (B^T B + \lambda I)^{-1} B \quad (6)$$

This is a slightly more greedy algorithm, but not as greedy as Broyden's first method (often called "Good Broyden") where  $A = S$ . (For mixing Broyden's first method has, to date, failed.) This algorithm can be considered as an adaptive algorithm, since if  $|S| \gg |Y|$  it becomes Broyden's first method while if  $|S| \ll |Y|$  it becomes Broyden's second method which is what is used in MSEC1 and MSEC3. In the earlier versions the constant  $\alpha = 1$ , with later version this is adjusted inside the code.

For the new algorithms where atomic positions are moved at the same time as the densities, all that needs to change is the definitions of  $s_{n-1}$  and  $y_{n-1}$  which now become:

$$s_{n-1} = (\rho_n - \rho_{n-1}, x_n - x_{n-1}) \quad ; \quad y_{n-1} = (R_n - R_{n-1}, g_n - g_{n-1}) \quad (7)$$

where  $x_n$  are the atomic positions, and  $g_n$  the (Pulay corrected) forces.

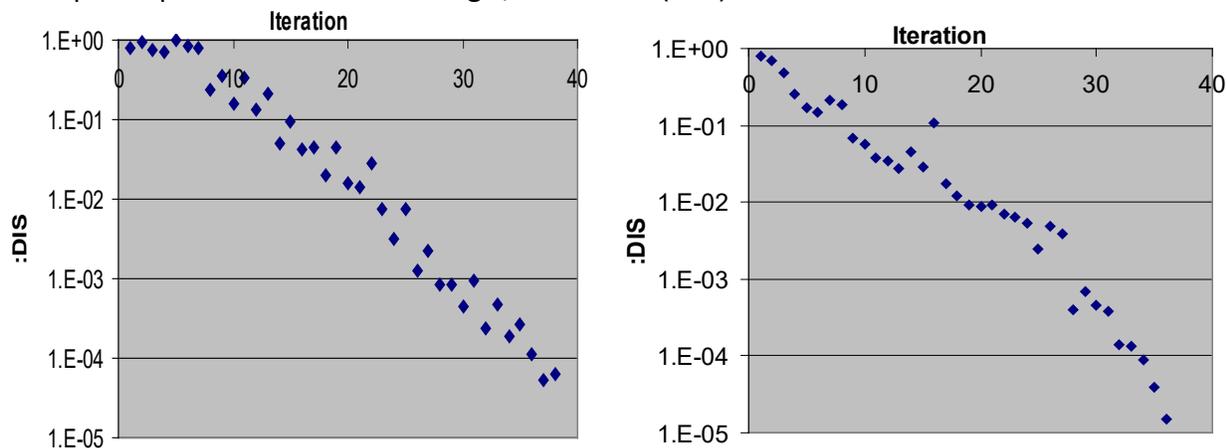
To understand how this can work, realize that the Jacobian used will now contain information not just on how the residue  $R_n$  changes as a function of density  $\rho_n$ , but also how the gradients (and Pulay corrections) change with density, the density changes with atomic positions as well as how the gradients change with atomic positions. Hence in effect it has a higher-order interpolation of the charge density for changes in the positions (higher-order than the currently very effective `clminterpol` code).

This can be a better method than using a double-loop since we are solving a variational problem even though we do not have a variational energy – the energy output by WIEN2k will converge to the true energy but only at the fixed point (self-consistent) solution. For fixed atomic positions we are in effect solving a constrained minimization, and the convexity of this (in effect how nice is the valley around the minimum) can be relatively poor. By opening up the problem so the atoms can move the problem can only become more convex, and consequently easier to solve.

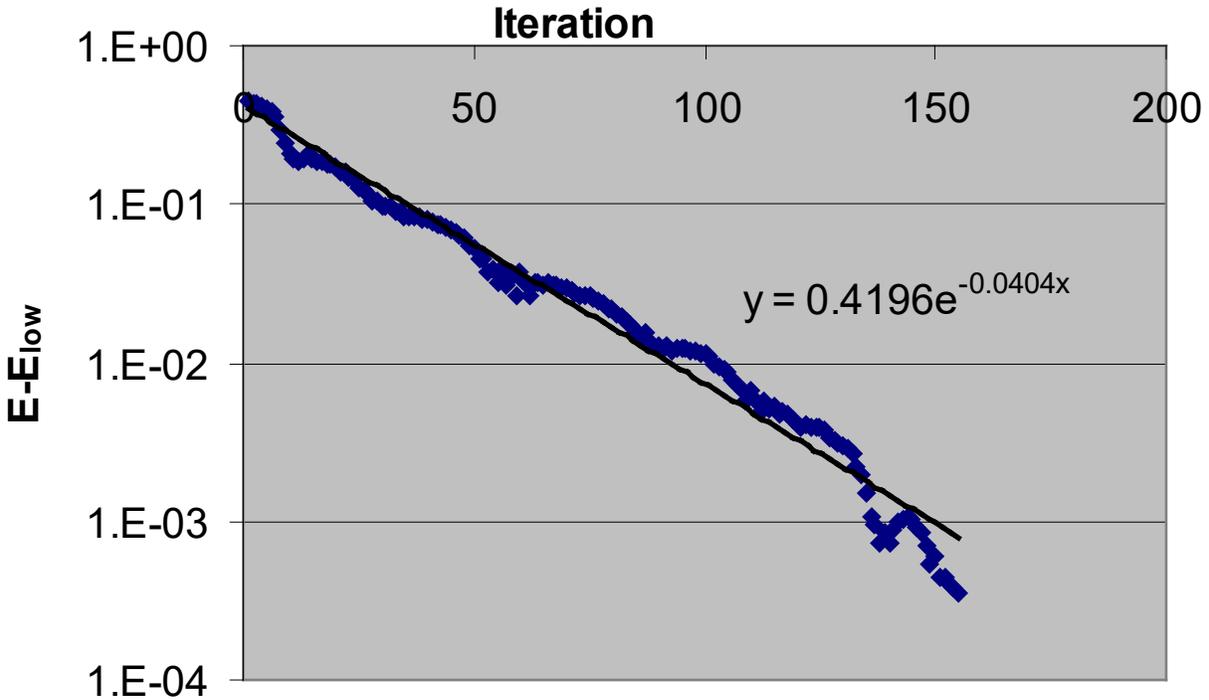
Of course, with a code such as WIEN2k this is only true if we do not have ghost-bands. Fortunately the change in how linearization energies are determined in version 10.1 has been a major improvement in this (thanks to Peter Blaha). In addition, because we have an implicit density extrapolation as mentioned above this is less likely. *Caveat: ghost-bands can always occur, so this cannot be guaranteed.*

The MSR1a code heads variationally downhill in energy, simultaneously making the densities converge and the forces smaller. Sometimes it will appear to be going uphill in energy or with the forces, but if you look carefully this is generally after some large displacements of the atoms and what is being adjusted in the next iterations is the densities. Like all Quasi-Newton methods, it is largely scale invariant so does not depend so much upon how the problem is formulated. That said, it is also known that the more tightly bunched are the eigenvalues of H, the faster will be the convergence. To improve this, there have been some changes in the variables used within the mixing to more appropriately take account of the effect of the multiplicity of the atoms as well as the degeneracy of the (hkl) plane waves and to use the same norm for the L=0 density within the muffin tins as that used for  $L \neq 0$ ; more details are given later. Another subtle change (in MSR1 only) in the earlier version was the introduction of a method for dynamically choosing the regularization parameter, but this has now been removed. An illustration of the convergence for MSEC3 is given below on the lower left for a relatively bad problem, a Ni (111) surface, which shows linear convergence; for well posed problems there is superlinear convergence. This can just about be seen for the problem on the lower right, a reduced TiO<sub>2</sub> c2x2 surface. (For reference, in both cases the initial densities were from dstart, i.e. a superimposition of isolated atomic densities.)

Similar to the older MSEC1, the convergence depends very weakly upon the mixing greed or the problem size (number of atoms) and is in the range of 10-50 self-consistent iterations, 10 for a well-posed problem such as bulk MgO, 40 for a Ni (111) surface or similar.



As an example of how MSR1a behaves, shown below is a plot for a relatively large problem of the change in energy referenced to the minimum as a function of energy (in Rydbergs). It shows a reasonable linear convergence with perhaps some superlinear convergence at the end although this is unclear.



One point to be careful about is the forces that are reported by MSR1a, as well as the path that the algorithm takes. It does *not* track along the Born-Oppenheimer surface, i.e. the surface of lowest energy for a given set of positions (which is what PORT does). It tracks on a different surface. Consider the “Surface of Zero Force”, i.e. the surface where for a given set of positions the forces on the atoms are zero, which will in general not be an electronic ground state. The algorithm tracks in between this surface and the Born-Oppenheimer, and the forces reported will in general be smaller (sometimes much smaller) than those of the Born-Oppenheimer surface. The two surfaces meet at the true minimum solution.

As a consequence you should always do a standard scf iteration to converge after MSR1a to check that you are close enough to the true solution. In addition, you should always converge with MSR1a for much smaller forces than one would normally use in PORT, coupled with a moderately tight charge convergence (e.g. `-cc 0.00025`) for a fully converged final result. That said, often it is not needed to converge to such a high level, for instance to test if a particular structure is feasible.

## Parallel Atomic Minimization Algorithms (a.k.a. the Energizer Bunny)

The MSR1a, MSECa algorithms are strange beasts. They are surprisingly stable, without extreme :DIS divergences and I have never seen ghost-bands with them (except with `-it`, but these are manageable in most cases tested so far and seem to have been patched by a change in `lapw1` and an improvement to `vec2pratt`). The new `-in1ef` option (default on) is particularly useful for them. While in principle they can use more accurate position and gradient information, they in fact work almost as well without this.

To use them:

- a) Do a preliminary scf convergence to something like `-cc 0.1 -ec 0.25 -fc 20` (I do not recommend a full convergence).
- b) Do "x pairhess", since the information in case.inM is used (both the constraints and trust region number). If you forget, the program will do this for you.
- c) If you have a badly posed problem, edit the atomic trust region in case.inM to something like 0.1. For a well posed problem the default of 0.2 is fine.
- d) Decide upon some convergence level, e.g. `-cc 0.001 -ec 0.001 -fc 1`, but be careful. How to determine when the algorithm has converged is tricky, and I do not have a full solution.
- e) Launch, and be patient since it can take > 100 or more iterations to converge, but this is still faster (1.5 to 2 times, perhaps better) than PORT in many cases. A nickname of "the Energizer Bunny" for these algorithms is appropriate, as they keep going and going downhill in energy.
- f) Run a final, conventional scf convergence after it has converged to "some level", and check that the forces are OK. (This is done automatically in the scripts.)

To date both full diagonalization and iterative diagonalization with `-it -vec2pratt -noHinv` have been fairly well tested. It is *strongly* recommended that `-vec2pratt` be used with the iterative diagonalization, it is much better.

Also, ensure that you have small enough R0 values for the atoms; use `lstart` and look for warnings. This can also lead to problems. Finally, ensure that you have a "good" case.in1. I suggest running `lstart` then editing case.in1\_st for the appropriate RKMAX, energy and number of bands and replacing old versions. Small steps in the linearization energy search turn out to be rather important.

After running the algorithm, it is safest to remove the precise positions at the end, for instance do "x patchsymm" then copy case.struct\_new to case.struct. This will ensure that you have a consistent structure file with atomic positions in the normal range of 0-1.0 .

**Note:** The algorithm uses a Trust Region (see later) to limit steps. The size of the step is printed out in the line where the GREED is also printed (use `grep -e :MIX`). This might become small for three reasons:

- 1) The code reads the first line of case.inM (if it exists) and uses  $\frac{1}{4}$  of the trust radius (default 0.35 a.u., damped by the value of :DIS to prevent the algorithm being too greedy far from the Born-Oppenheimer surface) as the largest step to take, and will reduce it if it is larger than this. (If case.inM is not present it will automatically execute pairhess to create it.)
- 2) It will only take a fraction of a full PRATT step.
- 3) The requested step would have brought atoms within 0.025 a.u. of each other, i.e. touching spheres.

Only 3) is a real concern, and if it occurs repeatedly you probably need to reduce the RMTs and use `clminter`; the others can be ignored

I strongly suggest MSR1a over MSECa, it appears to be much, much better – MSECa may fail to properly minimize and stagnates near saddle points.

The algorithms find a fixed point of  $(\rho - F(\rho), g)$  as a function of  $(\rho, x)$ , where

$\rho$	Electron Density
$F(\rho)$	SCF mapping of the density, i.e. CLMNEW, RHONEW
$g$	Gradients using :FOR
$x$	Atomic positions

For this to work, the gradient  $g$  has to approach the true gradient fairly quickly as the density converges, and it appears that it does. Beyond extending the prior  $S, Y$  information as well as the previous density and residue vectors to include  $x$  and  $g$ , almost nothing is different. For  $x$  and  $g$  “natural” units of a.u. and mRyd/a.u. appear to be correct.

It is *very important* to recognize that these algorithms are fundamentally different from MSEC3/PORT. For a well posed problem the two give the same result; for a badly posed problem they probably do not. These algorithms are sensitive to the convexity and global nature of the DFT problem being analyzed, and appear to be essentially variational in character (which in itself is a very interesting observation). So far they have been tested with PBE-GGA and PBE0 as well as LDA+U (see the later section on Ghost Forces).

The factor of 1.5-3.0 speed increase is not as good as I would like, and is going to be problem dependent. Some preliminary indications are that systems with soft modes (e.g. hydroxide) are a bit problematic, although these are also not so easy with PORT. Almost certainly there are some subtle noise issues in the current numerics of WIEN2k, not bugs but fundamental limits of the numerical algorithms (perhaps the differentiations/integrations/fits). I am hopeful that the speed will increase in the future.

One caveat: it is not impossible for the algorithm to find a saddle point, rather than going towards a real minimum. Whether this can in fact occur is currently unclear. (No precautions have been taken to force the Jacobian to be positive definite as is done for standard minimization codes.) To date it has not happened in tests with MSR1a. (If you start near a saddle point the progress may be very slow for a while.)

The algorithm can work with a badly posed problem, for instance a metallic surface with only a few k-point sampling of the Brillouin Zone and a small temperature factor (or tetrahedron mode). Note: I do not recommend the use of TEMP for any minimization (PORT as well) as it is not variationally correct, instead TEMPS should be used. For this type of case you can reduce the atomic trust region in case.inM down to perhaps 0.1 and use a GREED of 0.1, or reduce the Trust-Region using “TRUST 0.5” (see later). This, or some variant on this should work although in the long run it is probably better to use, for instance, more k-points.

Similar to a more conventional minimization using PORT, it is recommended that you do not use the largest RKMAX for initial minimizations starting far from a solution, rather a “good enough”

value. Again the same, you need to ensure that the muffin tins are far enough apart as otherwise the calculation will get stuck.

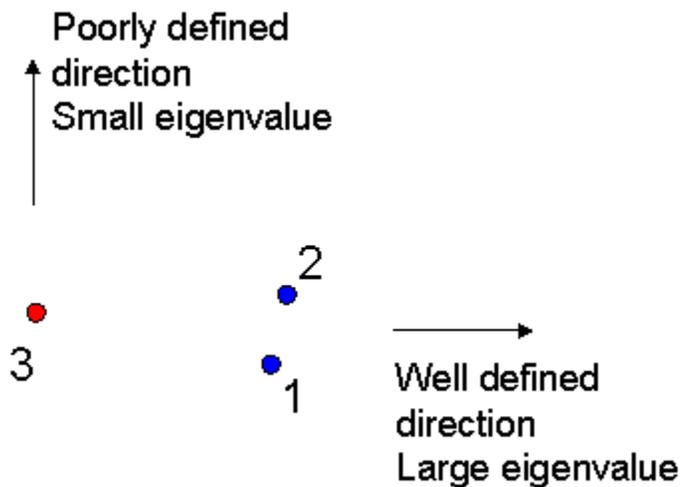
Finally, a strong statement and perhaps the most important use of MSR1a. If a particular functional/model combination fails to converge or converges badly with MSR1a, then it is > 98% probable that the combination is inappropriate. This follows from the fact that the algorithm appears to be variational, so a failure to converge or poor convergence implies weak to no convexity of the numerical method, i.e. it does not obey the KS variational principle. The most obvious case where this can occur is problems which are badly posed, for instance a metal without enough k-points for the Brillouin Zone sampling.

## Regularization

Regularization comes in when the algorithm calculates the vector  $b$  which satisfies (in a regularized fashion

$$(Y^T A)b = A^T F \tag{10}$$

where  $F$  is the current residue vector and  $A$  has been previously defined. What the algorithm does is used a SVD to regularize the matrix  $Y^T A$  (which is not symmetric) and from this calculate a pseudo-inverse. We need to use a regularization as otherwise the inverse of  $B$  may inflate eigenvalues/vectors which are mainly noise, leading to large steps in bad directions. To illustrate this, consider the diagram below where we are at point 3 and have prior information about points 1 and 2.



As marked, along the x axis the change (implicitly inside  $B$ ) is well defined, and the eigenvalue will be large; along the y axis it is less well defined and the eigenvalue will be smaller. When we solve for  $x_{\text{Reg}}$  above the small eigenvalue can have a large effect which is highly undesirable.

Therefore the term  $\lambda I$  is added (called a Tikonov regularization, similar to a Wiener filter), see for instance [http://en.wikipedia.org/wiki/Tikhonov\\_regularization](http://en.wikipedia.org/wiki/Tikhonov_regularization).

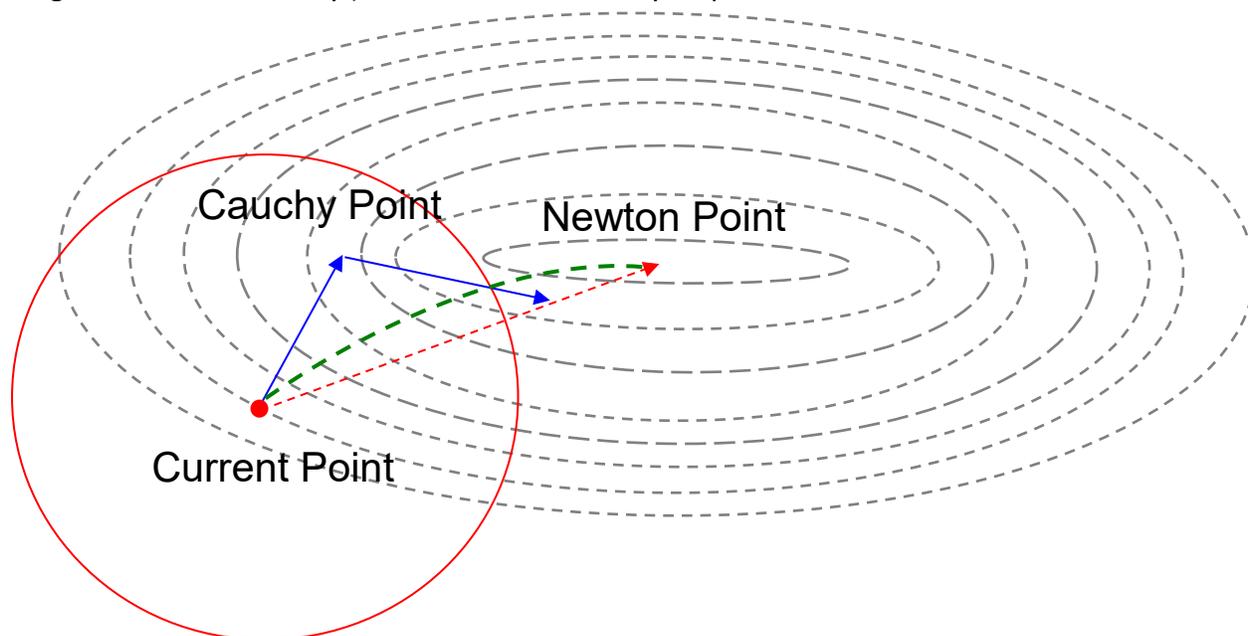
The current version uses a fixed term for lambda which can be adjusted.

## Trust Region Control

New with version 5.3 is a Trust Region control for overlarge steps. Taking  $r_i$  as the residual and  $p_i$  as some step, the approach is to use as a model the function

$$T = \frac{1}{2} (r_i - H^{-1}p_i)^2 \quad (12)$$

then, if the step requested is too large chose some appropriate step that minimizes in some sense  $T$ . The algorithm will first test a full Newton step; if this is too large or would lead to touching spheres it will move to a the Levenburg-Marquardt step for MSR1 (not implemented for MSEC3), then the reduced Newton step d) above, if this fails then c), b) and finally a) in order. This is illustrated below where (for simplicity) only the Cauchy Point is shown, with the red circle the Trust Radius, blue the double dogleg path and Green the Levenburg-Marquardt step. (Note: this diagram is schematic only.) In the code the Cauchy step is never used.



One additional, related change is that the algorithm now automatically reduces the atomic trust-region size if the structure contains atoms with d or f electrons. This appears to work well to improve stability for some metals which were harder to optimize with version 4.1. As a further step you can make the Trust Region even tighter by including at the end of case.inm a line "TRUST 0.5", or even "TRUST 0.25". Be aware that in the later case the algorithm may not properly converge because the Wolfe condition is being violated (i.e. too small steps to obtain meaningful new information) or it may be caught in a trap (a local minimum of the residue rather than a true root. This method is probably better than reducing the mixing GREED below 0.05. However, if you really need a small value here it probably means that there is something wrong with the physical model being used.

The code will output what it uses at every step, for instance (grep -e :MIX)

```
:MIX : MSR1a REGULARIZATION: 2.91E-02 GREED: 0.098 LMStep 0.338
```

```
:MIX : MSR1a REGULARIZATION: 2.54E-02 GREED: 0.200 Newton 0.877
```

```
:MIX : MSR1a REGULARIZATION: 3.07E-02 GREED: 0.200 Newton 1.000
```

For the Levenburg-Marquardt step the number reported is the absolute magnitude of the step relative to that of a full MSR1 step. In other cases the number is how far along the particular path the step is.

## Charge Renormalization in MSR1a

The standard renormalization of the charge density in Wien2k is to rescale the plane wave density so the total is correct. In a normal run this is appropriate for the input density, as any lost charge is due to poorly confined core densities so belongs in the plane waves. In early versions the same method was used in MSR1a for the density output from the algorithm.

This is not really correct or optimal for MSR1a. Any loss of density is due to inadequate interpolation of the effects of changing the positions of the atoms, which will transfer parts of the density from within the RMTs to the plane waves (or vica-versa). To a good approximation the core density will not depend upon the positions of the atoms. From 5.3 the charge renormalization is done with the total valence density (i.e. the density minus the core density), which is better.

## What are the Forces in MSR1a?

One source of confusion with MSR1a is that the Forces may appear to oscillate, and the atomic positions may move backwards and forwards at some stages. However, this is not something to be concerned about, and trying to adjust the behavior to avoid or eliminate this is inappropriate.

The key point is that the forces in MSR1a (and in Wien2k in general) are identical to the true forces **only** when the density is fully converged. When the density is not fully converged they are different, and are more correctly referred to as the **non self-consistent forces** or **pseudo-forces**. The algorithm is based upon the idea that these **pseudo-forces** converge to the true forces as the solution is approached, so it is rigorous to use their value within a fixed-point algorithm.

In some systems such as bulk magnesium oxide, the difference between the pseudo-forces and the true forces is small. In others such a materials with transition elements the difference can be quite large. The forces can then appear to oscillate, but this is a consequence of how well (or badly) the electron density is being treated.

To reduce these oscillations, the physical model of the electrons has to be improved, which may not be so simple. Often this involves using more k-points, avoiding shifted k-meshes and reducing RMTs a little. It is also desirable to stabilize the pseudo-charge of the plane waves within the muffin tins, although how to do this is currently not clear.

Reducing the effect of the non-self consistent contributions is an unsolved problem, and probably depends more upon the internal working of Wien2k including calculation of exchange-correlation potentials than anything else.

## Ghost Bands and Ghost Forces

Primarily with iterative modes in WIEN2k one can have Ghost Bands. This can occur if there has been a bad choice of linearization energies in case.in1, but this is largely reduced by the `-in1ef` switch in the more recent releases. In the iterative modes it can also occur if the linearization energies of the states used from the previous iteration are bad approximations to those for the current iteration. When atoms are moving this is more likely to occur.

Something (undesirable) new is what I am calling "Ghost Forces", for want of a better name. What happens is that the force for some atoms suddenly becomes very large, sometimes  $> 1000$  mRyd/au without anything apparently wrong with the density or energies. This only occurs with the iterative mode. The problem appears to be related to breakdown of orthogonality of the plane waves associated with the LO's in `lopw.f` and a patch is being tested. A patch to `lopw.f` appears to solve this problem in most cases, but it is not impossible that it can still occur in some cases.

With normal modes (e.g. PBE) the current mixer appears to be sufficiently robust that it can handle this problem in most cases tested so far. At the time of writing this (August 5, 2010) the switch `-noHinv` -it appears to be best (`-vec2pratt` is OK, maybe not so good). The mode where a single-precision approximation to the inverse of H inside `lapw1` is stored (the default) does not appear to be cost effective; it reduces the time per scf iteration but this is offset by the number of extra cycles needed, and it does not converge as well. As a caveat, the performance of the different modes with a changed `lopw.f` has not yet been benchmarked. A few non-iterative cycles should probably be run at the end of a MSR1a calculation just to ensure that nothing strange has happened.

For weakly correlated systems, e.g. the Ti d-electrons in  $\text{TiO}_2$  or  $\text{SrTiO}_3$  compounds/surfaces the Atoms mode appears to work fine in iterative mode using PBE0 (-it `-noHinv` `-vec2pratt`). However, for strongly correlated systems such as a NiO surface using GGA+U or PBE0 the algorithm can fail due to the ghost forces if `lapw1` has not been patched.

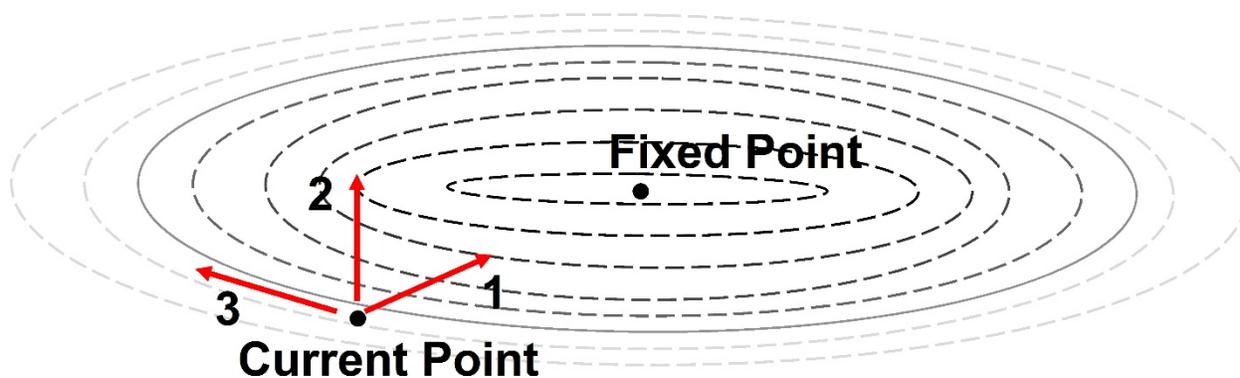
One change with the Mango release is that if there is a non-catastrophic QTL-B problem (ghost-band), the code will reduce the trust region. The code also reduces the trust region when it detects large fluctuations in the potential.

## Dealing with Hard Problems

In some cases the convergence of the mixer is poor, independent of the mode used (MSEC3, MSR1), and the behavior of the atom movement algorithm MSR1a can appear to be bad with severe oscillations. While this may worry you, *in general it should not*. Convergence problems are normally because your problem is ill-conditioned. If you encounter very poor convergence, you should analyze whether you have posed the problem correctly in terms of the physics; in many cases changing this solves all problems.

To expand a little, all methods both in Wien2k as well as in the literature use some variant of an expansion of the information from previous steps to form a Jacobian. Which directions are used to collect this information comes from the residue, the difference between the input density at the start of an iteration and the new one after, the Pratt step. What matters is the properties of this Pratt step.

For problems which converge quickly this step is very strongly downhill; for problems which converge badly it is not. For the “good” problems which converge well the algorithm is rapidly obtaining information about the main directions towards the self-consistent solution. For bad problems it is obtaining less information about where the solution is, rather information about less useful directions. This is illustrated below: if the Pratt step is along 1 the fixed-point solution will be found quickly; if along 2 it will still be found but a bit slowly; for 3 it will be slow.



You can estimate what you have for a given problem by looking at the angles between the Broyden and Pratt steps which are printed in case.scfm. If these are near 90 degrees or are larger than this expect the convergence to be slow.

Changing from TETRA to TEMPS (or perhaps TEMP) is known, empirically, to make convergence better by improving the Pratt step – exactly how is not known. Similarly increasing the number of k-points for a metallic system seems to help, although here the effect may be a reduction of telegraph noise in TETRA. Reducing the upper bound of the GREED can also be useful, perhaps 0.1. Reducing more than this is normally a very, very bad idea. The new **STIFF** mode may handle many of these problems.

## Cases where MSR1a works well

1. Volume relaxation: forces converge cleanly as the density converges.
2. Change of RMTs by a small amount – similar to 1. Caveat, when changing RMTs by a lot it may be necessary to create a new density, as `clminter` is not good in such cases.
3. Change of RKMAX, e.g. after a minimization – similar to 1.
4. Change of functional after having done a previous run, e.g. moving to PBE0. Recommendation: use `clmextrapol` with appropriate lattice changes to accommodate any volume differences, in general it seems that a homogeneous expansion/contraction is the leading change.
5. Minimizations where a good scaling of the initial estimate in pairness is not known, or when there are large changes in the Hessian during the minimization, e.g. bonding changes. In such a case PORT can take some time to lose the old Hessian information (a number of fully converged outer loops) whereas MSR1a will lose this old information in a few iterations, the number of memory steps used.

## Older Version Notes

### Major Changes with Version 4.1

- 1) The old Broyden method (qmix5) has been retired.
- 2) A variant of the Pratt mode (PRAT0) uses the NFL/Trust-region control of MSEC1, so PRAT0 0.5 should be safe (albeit slow).
- 3) The case.clmsum/up/dn files have one digit higher precision by use of ES rather than E for output. This *does not* make them any bigger.
- 4) Relevant parts of mixer.F have been vectorized/converted to BLAS. Redundant parts of the code have been eliminated and many of the subroutines used simplified.
- 5) The prior MSEC1/MSEC2 is still in the code, in subroutines which all have a “7” in them. Newer versions have an “8” in them. The behavior of MSEC1/MSEC2 should be identical to before, with minor changes due to 3) and 4).
- 6) Newer versions contain vectorized/BLAS calls, more for consistency and accuracy than anything else. Note: W2kinit.F sets high-precision for the vectorization calls. This should reduce/avoid issues with ifort reduced-accuracy optimizations.
- 7) MSEC3/MSEC4 are new versions of MSEC1/MSEC2. They have a consistent L2 scaling of all the variables including atom multiplicity as well as the plane wave degeneracy. They are slightly better in some cases than the old versions, although with  $-in1ef$  it is hard to say. For certain, they should behave the same for different problems (equally good/bad).
- 8) The algorithms MSR1, MSR2 are new algorithms based upon a rank-one multiseccant update; MSR2 uses only LM=0 similar to MSEC2. The MSEC1 and MSR1 families can both be represented in terms of a Jacobian in equations (3) & (5) above. If  $\beta = 1$  MSR1 is similar to a multiseccant version of the symmetric rank-one update (SR1). This algorithm is slightly more greedy than MSEC3, and (more importantly) is more appropriate for optimization problems. For methods using simultaneous atom minimization it is much better. Whether MSR1 is better than MSEC3 probably depends more upon the class of problem, and how convex is the physical model/numerics. At present MSR1 appears to be slightly faster and equally robust in most cases.
- 9) The algorithms MSR1a, MSR2a, MSECa, MSECb do a simultaneous “mixing” of densities and atomic positions. MSR2a is a LM=0 only version of MSR1a (untested), similarly MSECb. They add the gradients (FOR required in case.in2) and atomic positions. See later for notes on MSR1a. MSECa and MSECb are not recommended.
- 10) A new algorithm is used to choose the regularization parameter within a certain range for MSR1 and MSR1a via a Generalized Cross-Validation, which allows the algorithm to work better both for well-posed problems as well as poorly posed ones. It is not used by default for MSEC3 as a fixed regularization appears to be just as good in this case.
- 11) The maximum unpredicted greed in the early stages when :DIS is large has been raised, as this was leading to bad convergence as this violated Wolfe conditions.
- 12) All the new algorithms use a different method of handling the inverse in the Jacobian which gives an estimate of the conditioning (COND) defined as the condition number of the matrix  $(Y^T A + \lambda I)$  multiplied by the regularization term  $\lambda$  and divided by the number of memory steps.

This is useful information; the smaller the condition number, the more convex the problem (probably on a log scale). This is now output in the :DIRM line.

- 13) All the new algorithms give scaling for the relevant PW/DM/Atoms in case.scfm. The Atom modes also print out the positions and displacements (:APOSXXX for grep).
- 14) The term "MIXING SCHEME WITH XXX" in the old version has been replaced by "MIXING GREED XXX" in the hope that users will start to understand that XXX is *not* what most people think it is. Just as being too greedy is bad, being not greedy enough is also bad – the algorithm starves to death. (More technically, it will not come close to satisfying Wolfe conditions. Note: Wolfe conditions for fixed-point iterations do not appear to be documented in the literature, but everything points to them being real – a paper for a mathematician.)
- 15) A command "grep -e :DIR -e GREED -e :FRMS -e :ENE -e :CHARG -e PRATT -e :DIS -e "MIXING SC" -e PLANE \*.scf \$1 | tail -40" will catch most useful output, e.g. giving

```
:DIS : CHARGE DISTANCE ( 0.0001591 for atom 16 spin 1) 0.0000583
:PLANE: INTERSTITIAL TOTAL 7.77491 DISTAN 5.136E-03 %
:CHARG: CLM CHARGE /ATOM 28.25380 DISTAN 3.584E-04 %
:DIRM : MEMORY 6/8 RESCALE 2.61 RED 0.597 PRED 0.781 NEXT 0.691 COND 1.39E-01
:DIRP : |BROYD|= 6.586E-04 |PRATT|= 1.510E-04 ANGLE= 18.7 DEGREES
:DIRB : |BROYD|= 1.220E-03 |PRATT|= 3.694E-04 ANGLE= 15.5 DEGREES
:MIX : MSEC3 REGULARIZATION 5.00E-05 GREED 0.145 Newton 1.000
:ENE : ***** TOTAL ENERGY IN Ry = -116780.86459533
```

The scaling of :PLANE and :CHARG is different with the new modes, only the % values are comparable but these not completely because a L2 norm is now used. The "REGULARIZATION" is the value of  $\lambda$  used, which by default varies (see later). For an atom minimization the output will look more like

```
:DIS : CHARGE DISTANCE ( 0.0007947 for atom 17 spin 1) 0.0002363
:PLANE: INTERSTITIAL TOTAL 7.77798 DISTAN 1.110E-01 %
:CHARG: CLM CHARGE /ATOM 28.25379 DISTAN 1.303E-03 %
:DIRM : MEMORY 8/8 RESCALES 1.06 0.93 RED 1.276 PRED 1.000 NEXT 1.000 COND 5.50E-02
:DIRA : |BROYD|= 5.444E-02 |PRATT|= 7.006E-04 ANGLE= 69.7 DEGREES
:DIRP : |BROYD|= 4.652E-02 |PRATT|= 7.962E-04 ANGLE= 85.8 DEGREES
:DIRB : |BROYD|= 5.000E-02 |PRATT|= 1.087E-03 ANGLE= 86.8 DEGREES, STEP= 0.51077
:FRMS (mRyd/au) 1.313 :DRMS (au) 7.067E-03
:MIX : MSR1a REGULARIZATION 3.81E-04 GREED 0.087 Newton 1.000
```

Here :DIRA is the atomic step, two scales are show (one for the PW, one for the atoms) and :FRMS is the RMS forces with :DRMS the RMS atomic displacement of this cycle.

For orbital potentials, density matrices an additional "DIRD" term is show and the 2<sup>nd</sup> scaler is that for these. For orbital potentials and atoms there are three scalings shown (not tested).

- 16) The atomic mixing outputs more precise atomic positions (more decimal points) at the end of the case.struct file.

## Major Changes with Version 5.3

Version 5.3 is a fairly substantial upgrade of the mixer. While the core algorithms (MSEC3/MSR1/MSR1a) have not changed, a much improved method of controlling the overall step using a trust-region method has been added, and some simple limits added (as defaults) to reduce the algorithm greed for problems involving d or f electrons (which tend to converge poorly in Wien2k due to Fermi-surface sloshing). For simple cases with the standard non-optimizing algorithms (MSEC3/MSR1) in most cases one will only see a small change, reduction of overly aggressive steps. (The trust region is better than MSR1, and since this algorithm appears to be better in most cases than MSEC3 it is now the recommended option and the main target for development.) Particularly for metals involving d or f electrons the performance of MSR1a should be much improved, although it does not compare to MSR1a for insulators which, for reasons associated with the physics of the problem, will almost always converge faster. In more detail:

- 1) A trust-region method has been added to control overlarge steps (new in 5.3). This, combined with other changes improves the stability for what were hard cases with the initial version
- 2) The algorithms are now fully controlled by the implicit trust region of the greed in all cases. In effect, the total atom step and the total step cannot exceed  $a \cdot \text{GREED}$  where  $a$  is a constant (set internally for the two cases) and GREED is the locally adapted greed as output in case.scfm (`grep -e :MIX *.scf`).
- 3) The code now understands when the user is employing the mBJ potential and will allow mixing of vresp to be controlled by the GREED in case.inm\_vresp. (Previously it was always 1.0.) A value of 0.5 in case.inm\_vresp is reasonable. Note that in other cases mixing of vresp is over-ridden and a value of 1.0 with PRATT mixing is used. In all cases the density in vresp is not normalized.
- 4) A different method is used to normalize the density in MSR1a. For version 4.1 any density lost was added to the plane waves; with 5.3 it is added to the valence density instead.
- 5) Very small forces are now trapping; in principle they could lead to instabilities even with some simple cases (e.g. an Al surface) in 4.1
- 6) :DIS is now more correct as an L2. The numbers *are different*, but in most cases this will not be apparent.
- 7) :FRMS is now more correct.
- 8) Sudden increases in the total step are now trapped.

## Major Changes with Version 6.0

Version 6.0 has large changes in many details of the algorithm, the new version hopefully being both more convergent and more robust. Some issues in the prior version which would not always converge well soft modes appears to have been solved.

A detailed description of this algorithm can be found in *Fixed-Point Optimization of Atoms and Density in DFT*. Marks, L.D., Journal of Chemical Theory and Computation, 2013. 9(6), 2786-2800.

## Major changes with Version 7.1.2

Version 7.1.2 will appear essentially the same to the user, but in operation has some significant differences. The earlier version used mainly the Greed (sometimes inaccurately called the mixing factor) as an implicit trust region control, i.e. it was kept small when the linear model of the expansion of the Jacobian was not that good. An explicit trust region control via the total step length and the movement of the atoms was more a secondary safety constraint.

While this can work, it turned out to be weak for cases where there are soft modes. In these cases the Greed needs to be larger even if the algorithm is not making substantial improvements. To handle this, version 7.1.2 uses larger values of the Greed and the explicit trust region much more and also has an additional explicit trust region control for the total charge change within any of the muffin tins. As such it works much better for soft modes and seems to still work very well for other standard problems.

One addition in Version 7.1.2 is (with MSR1a) a printout via a different method of an estimate of how far the current positions are from the minimum. While this is not an exact metric, since it uses a different algorithm it works as a good check. For a true minimum the density differences, forces as well as this different should all be small; there are cases where the density differences and forces seem to be small but from this other metric it is apparent that there is still some distance to the minimum – a characteristic of soft modes.

## Major Changes with Version 8.0

Mixer 8.0 was a minor update to release 7.1.2, with a few small improvements in some of the controls, but nothing very major.