Geometry Optimization: Ver. 1.0

Taisuke Ozaki, RCIS, JAIST

August 16, 2007

1 Newton method

The total energy E of a system can be expanded by the Taylor series with respect to atomic coordinates $\{x_i\}$ around E_0 with $\{x_i^{(0)}\}$ as follows:

$$E = E_0 + \sum_i^{3N} \left(\frac{\partial E}{\partial x_i}\right)_0 (x_i - x_i^{(0)}) + \frac{1}{2}\sum_{i,j}^{3N} \left(\frac{\partial^2 E}{\partial x_i \partial x_j}\right)_0 (x_i - x_i^{(0)})(x_j - x_j^{(0)}) + \cdots, \tag{1}$$

where the derivatives $()_0$ mean the partial derivatives at $\{x_i^{(0)}\}$, and N is the number of atoms. By differentiating Eq. (1) with respect to x_k, to the second order we have

$$\frac{\partial E}{\partial x_k} = \left(\frac{\partial E}{\partial x_k}\right)_0 + \sum_i^{3N} \left(\frac{\partial^2 E}{\partial x_k \partial x_i}\right)_0 (x_i - x_i^{(0)}). \tag{2}$$

In case the coordinates $\{x_i\}$ give a local minimum, assuming $\frac{\partial E}{\partial x_k} = 0$, we have the following matrix equation:

$$\begin{pmatrix} (\frac{\partial^2 E}{\partial x_1 \partial x_1})_0 & (\frac{\partial^2 E}{\partial x_1 \partial x_2})_0 & \cdots \\ (\frac{\partial^2 E}{\partial x_2 \partial x_1})_0 & (\frac{\partial^2 E}{\partial x_2 \partial x_2})_0 & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} (x_1 - x_1^{(0)}) \\ (x_2 - x_2^{(0)}) \\ \cdots \end{pmatrix} = - \begin{pmatrix} \left(\frac{\partial E}{\partial x_1}\right)_0 \\ \left(\frac{\partial E}{\partial x_2}\right)_0 \\ \cdots \end{pmatrix}. \tag{3}$$

The short notation is

$$H\Delta\mathbf{x} = -\mathbf{g}, \tag{4}$$

where the matrix consisting of the second derivatives in the left-hand side is called Hessian H. Using Eq. (4), $\{x_i\}$ can be updated by

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - (H^{(n)})^{-1}\mathbf{g}^{(n)}. \tag{5}$$

This is the well known Newton method.

2 RMM-DIIS

In the OpenMX, $\mathbf{x}^{(n)}$ and $\mathbf{g}^{(n)}$ in Eq. (5) are replaced by $\bar{\mathbf{x}}^{(n)}$ and $\bar{\mathbf{g}}^{(n)}$ given by the residual minimization method in the direct inversion of iterative subspace (RMM-DIIS) [1, 2] as follows:

$$\mathbf{x}^{(n+1)} = \bar{\mathbf{x}}^{(n)} - \alpha \ (H^{(n)})^{-1}\bar{\mathbf{g}}^{(n)}, \tag{6}$$

where α is a tuning parameter for acceleration of the convergence, which can be small (large) for a large (small) $\bar{\mathbf{g}}^{(n)}$. $\bar{\mathbf{g}}$ in the RMM-DIIS can be found by a linear combination of previous upto p-th gradients $\mathbf{g}$ as

$$\bar{\mathbf{g}}^{(n)} = \sum_{m=n-(p-1)}^{n} a_m \mathbf{g}^{(m)}, \tag{7}$$

where a_m is found by minimizing $\langle \bar{\mathbf{g}}^{(n)} | \bar{\mathbf{g}}^{(n)} \rangle$ with a constraint $\sum_{m=n-(p-1)}^{n} a_m = 1$. According to Lagrange's multiplier method, F is defined by

$$
\begin{aligned}
F &= \langle \bar{\mathbf{g}}^{(n)} | \bar{\mathbf{g}}^{(n)} \rangle - \lambda(1 - \sum_m^n a_m), \\
&= \sum_{m,m'} a_m a_{m'} \langle \mathbf{g}^{(m)} | \mathbf{g}^{(m')} \rangle - \lambda(1 - \sum_m^n a_m).
\end{aligned} \tag{8}
$$

Considering $\frac{\partial F}{\partial a_k} = 0$ and $\frac{\partial F}{\partial \lambda} = 0$, an optimum set of $\{a\}$ can be found by solving the following linear equation:

$$
\begin{pmatrix}
\langle \mathbf{g}^{(n-(p-1))} | \mathbf{g}^{(n-(p-1))} \rangle & \cdots & \cdots & 1 \\
\cdots & \cdots & \cdots & 1 \\
\cdots & \cdots & \langle \mathbf{g}^{(n)} | \mathbf{g}^{(n)} \rangle & \cdots \\
1 & 1 & \cdots & 0
\end{pmatrix}
\begin{pmatrix}
a_{(n-(p-1))} \\
a_{(n-(p-1)+1)} \\
. \\
\frac{1}{2}\lambda
\end{pmatrix}
=
\begin{pmatrix}
0 \\
0 \\
. \\
1
\end{pmatrix}. \tag{9}
$$

An optimum choice of $\bar{\mathbf{x}}^{(n)}$ may be obtained by the set of coefficients $\{a\}$ as

$$\bar{\mathbf{x}}^{(n)} = \sum_{m=n-(p-1)}^{n} a_m \mathbf{x}^{(m)}. \tag{10}$$

If the Hessian H is approximated by the unity I, Eq. (6) becomes

$$\mathbf{x}^{(n+1)} = \bar{\mathbf{x}}^{(n)} - \alpha \, \bar{\mathbf{g}}^{(n)}. \tag{11}$$

This scheme in the Cartesian coordinate has been implemented as 'DIIS' in OpenMX.

3 Broyden-Fletcher-Goldfarb-Shanno (BFGS) method

Define

$$
\begin{aligned}
\Delta \mathbf{g}^{(n)} &= \mathbf{g}^{(n)} - \mathbf{g}^{(n-1)}, \tag{12} \\
\Delta \mathbf{x}^{(n)} &= \mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}. \tag{13}
\end{aligned}
$$

Then, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [3] gives the following rank-2 update formula for $(H^{(n)})^{-1}$:

$$
\begin{aligned}
(H^{(n)})^{-1} &= (H^{(n-1)})^{-1} + \frac{\langle \Delta \mathbf{x}^{(n)} | \Delta \mathbf{g}^{(n)} \rangle + \langle \Delta \mathbf{g}^{(n)} | (H^{(n-1)})^{-1} | \Delta \mathbf{g}^{(n)} \rangle}{\left(\langle \Delta \mathbf{x}^{(n)} | \Delta \mathbf{g}^{(n)} \rangle \right)^2} | \Delta \mathbf{x}^{(n)} \rangle \langle \Delta \mathbf{x}^{(n)} | \\
&\quad - \frac{(H^{(n-1)})^{-1} | \Delta \mathbf{g}^{(n)} \rangle \langle \Delta \mathbf{x}^{(n)} | + | \Delta \mathbf{x}^{(n)} \rangle \langle \Delta \mathbf{g}^{(n)} | (H^{(n-1)})^{-1}}{\langle \Delta \mathbf{x}^{(n)} | \Delta \mathbf{g}^{(n)} \rangle},
\end{aligned} \tag{14}
$$

where $(H^{(0)})^{-1} = I$. An optimization scheme using Eq. (6) and the BFGS update formula for the inverse of an approximate Hessian matrix in the Cartesian coordinate has been implemented as 'BFGS' in OpenMX.

4 Rational function (RF) method

The BFGS update by Eq. (14) without any care gives an ill-conditioned approximate inverse of Hessian having negative eigenvalues in many cases. This leads to the optimization to saddle points rather than the optimization to a minimum. The rational function (RF) method [4] can avoid the situation in principle. Instead of Eq. (1), we may consider the following expression:

$$E \;=\; E_0 + \sum_i^{3N} \left(\frac{\partial E}{\partial x_i}\right)_0 (x_i - x_i^{(0)}) + \frac{1}{2}\sum_{i,j}^{3N}\left(\frac{\partial^2 E}{\partial x_i \partial x_j}\right)_0 (x_i - x_i^{(0)})(x_j - x_j^{(0)}) + \frac{1}{2}\lambda\sum_i^{3N}(x_i - x_i^{(0)})^2. \quad (15)$$

Then, the equation corresponding to Eq. (4) becomes

$$(H^{(n)} + \lambda I)\Delta \mathbf{x}^{(n)} = -\mathbf{g}^{(n)}. \quad (16)$$

Therefore, a large λ assures that $(H^{(n)} + \lambda I)$ is positive definite. If $\lambda^{(n)}(= -\lambda)$ is given by

$$\lambda^{(n)} = \langle \mathbf{g}^{(n)} | \mathbf{s}^{(n)} \rangle. \quad (17)$$

With Eq. (17), Eq. (16) may be equivalent to

$$\begin{pmatrix} H^{(n)} & \mathbf{g}^{(n)} \\ (\mathbf{g}^{(n)})^T & 0 \end{pmatrix}\begin{pmatrix} \Delta \mathbf{x}^{(n)} \\ 1 \end{pmatrix} = \lambda^{(n)}\begin{pmatrix} \Delta \mathbf{x}^{(n)} \\ 1 \end{pmatrix}, \quad (18)$$

where the size of the matrix in the left-hand side is $(3N + 1) \times (3N + 1)$, and called the augmented Hessian. The lowest eigenvalue of the eigenvalue problem defined by Eq. (18) may give an optimum choice for λ, and the corresponding eigenvector, the last component is scaled to 1, gives an optimization step $\Delta \mathbf{x}^{(n)}$. In Eq. (18), the approximate Hessian can be estimated by the following BFGS formula:

$$H^{(n)} \;=\; H^{(n-1)} + \frac{|\Delta \mathbf{g}^{(n)}\rangle\langle\Delta \mathbf{g}^{(n)}|}{\langle\Delta \mathbf{x}^{(n)}|\Delta \mathbf{g}^{(n)}\rangle} - \frac{H^{(n-1)}|\Delta \mathbf{x}^{(n)}\rangle\langle\Delta \mathbf{x}^{(n)}|H^{(n-1)}}{\langle\Delta \mathbf{x}^{(n)}|H^{(n-1)}|\Delta \mathbf{x}^{(n)}\rangle}, \quad (19)$$

where $H^{(0)} = I$. An optimization scheme using Eq. (18) and the BFGS update formula Eq. (19) in the Cartesian coordinate has been implemented as 'RF' in OpenMX.

5 Eigenvector following (EF) method

By diagonalizing the approximate Hessian given by Eq. (19), the ill-conditioned situation can be largely reduced [5]. The approximate Hessian is diagonalized as

$$E^{(n)} = (V^{(n)})^T H^{(n)} V^{(n)}, \quad (20)$$

where $E^{(n)}$ is a diagonal matrix of which diagonal parts are eigenvalues of $H^{(n)}$. If the eigenvalue of the approximate Hessian is smaller than a threshold (0.02 a.u. in OpenMX3.3), the eigenvalue is set to the threshold. The modification of eigenvalues gives a corrected matrix $E'^{(n)}$ instead of $E^{(n)}$. Then, we have the inverse of a corrected Hessian matrix $H'^{(n)}$ being a positive definite as

$$(H'^{(n)})^{-1} = V^{(n)}(E'^{(n)})^{-1}(V^{(n)})^T. \quad (21)$$

A optimization scheme using the inverse Eq. (21) in Eq. (6) in the Cartesian coordinate has been implemented as 'EF' in OpenMX.

In addition, there are two important prescriptions for the stable optimization: (1) If $\langle \Delta \mathbf{x}^{(n)} | \Delta \mathbf{g}^{(n)} \rangle$ is positive in the update of Hessian by Eq. (19), it is assured that the updated Hessian is positive definite. Therefore, if $\langle \Delta \mathbf{x}^{(n)} | \Delta \mathbf{g}^{(n)} \rangle$ is negative, the update should not be performed. (2) The maximum step should be always monitored, so that an erratic movement of atomic position can be avoided.

References

[1] P. Csaszar and P. Pulay, J. Mol. Struc. **114**, 31 (1984).

[2] F. Eckert, P. Pulay, and H.-J. Werner, J. Comp. Chem. **18**, 1473 (1997).

[3] C. G. Broyden, J. Inst. Math. Appl. **6**, 76 (1970); R. Fletcher, Comput. J. **13**, 317 (1970); D. Goldrarb, Math. Comp. **24**, 23 (1970); D. F. Shanno, Math. Comp. **24**, 647 (1970).

[4] A. Banerjee et al., J. Phys. Chem. **89**, 52 (1986).

[5] J. Baker, J. Comput. Chem. **7**, 385 (1986).