Algorithm Improvement for Finite-Temperature Structural Optimization Using Anharmonic Phonon Theory

Takumi Chida[1], Ryota Masuki[2]

[1] Tsuneyuki Laboratory, Department of Physics, Graduate School of Science, The University of Tokyo

[2] Arita Laboratory, Department of Applied Physics, Graduate School of Engineering, The University of Tokyo

September 27, 2024

About of Authors

- Takumi Chida

  He developed new first-principles electronic structure calculation methods using the transcorrelated method. In this study, he proposed a numerical optimization algorithm and performed calculations using test functions.

- Ryota Masuki

  He developed the SCPH-based structural optimization and implemented it in the ALAMODE package. In this study, he was involved in the discussion on the test calculations and the implementation of the new feature in the ALAMODE package.

1 Background

Many materials are known to exhibit changes in crystal structure with temperature, such as ferroelectrics, charge density waves, and martensitic transformations. These structural phase transitions are not only important for practical applications but are also of interest because various intriguing phenomena, such as anomalies in transport properties, are observed near phase transitions.

Traditionally, calculations of crystal structures using density functional theory (DFT) have been limited to zero temperature, making it difficult to calculate the temperature dependence of crystal structures. However, in recent years, by advancing the theory of anharmonic phonons, a theoretical method capable of calculating the temperature dependence of crystal structures has been developed. This method has been made available as a new feature in the open-source software for anharmonic phonon calculations, ALAMODE [1, 2, 3]. This method has been validated with simple perovskite materials like $BaTiO_3$ and, in principle, can be applied to a wide range of materials. However, with the current algorithm used for structural updates, convergence of the crystal structure is slow, making it practically difficult to apply to complex materials.

Therefore, this study aims to develop an algorithm that operates robustly in general cases and converges structures more efficiently, and to implement it in the ALAMODE program. RMM-DIIS[4] is a well-known method for accelerating convergence in function optimization, but this method searches for extrema of the function, which may lead to convergence to unstable, high-symmetry phases near saddle points of the potential surface during structural phase transitions. To address this issue, this study implemented the BFGS [5] method, which more robustly searches for minima, and methods combining these approaches for efficient convergence. The performance

of these methods was tested with cases focusing on structural phase transitions. Furthermore, efforts were made to implement these methods in the ALAMODE package to allow easy switching between various optimization methods in structural optimization.

2 Theory of Optimization Methods

In this section, we briefly introduce the Newton method that has been used in previous calculations, its variant, the BFGS method, another useful approach, the RMM-DIIS method, and its challenges, and finally the GDIIS method that is the focus of this study.

2.1 Newton Method

We consider the problem of optimizing a function $f(\boldsymbol{x})$ with respect to its argument. Let the value of $\boldsymbol{x}$ at the k-th step be $\boldsymbol{x}_k$, the gradient of the function f at that point be $\boldsymbol{g}(\boldsymbol{x}_k)$, and the Hessian be H_k^{-1}. The Newton method updates the structure as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H_k^{-1}\boldsymbol{g}(\boldsymbol{x}_k) \tag{2.1}$$

Currently, the ALAMODE package uses a method based on the Newton method. However, because calculating the free energy Hessian is difficult, an approximate estimate that is simpler to compute is used instead, which may result in significantly slower convergence along the soft mode directions where the potential surface flattens out.

2.2 RMM-DIIS Method

Instead of using a fixed, coarse approximation for the Hessian, a method that updates the approximation of the Hessian at each step using the gradient of the free energy can accelerate convergence. The RMM-DIIS (Residual Minimization Method with Direct Inversion in the Iterative Subspace) method is one such technique, which improves convergence by using the gradient history from past search points [4]. In RMM-DIIS, the update is performed using the past L search points as follows:

$$\boldsymbol{x}_{k+1} = \sum_{i=k-L+1}^{k} c_i \boldsymbol{x}_i, \qquad \sum_{i=k-L+1}^{k} c_i = 1 \tag{2.2}$$

The coefficients c_i are determined by minimizing

$$W_k = \frac{1}{2}|\tilde{\boldsymbol{g}}_{k+1}|^2 + \lambda\left(\sum_{i=k-L+1}^{k} c_i - 1\right), \quad \tilde{\boldsymbol{g}}_{k+1} = \sum_{i=k-L+1}^{k} c_i \boldsymbol{g}(\boldsymbol{x}_i) \tag{2.3}$$

This is because the gradient at a point $\boldsymbol{x}$ close enough to an extremum $\boldsymbol{x}^*$ can be approximated as

$$\boldsymbol{g}(\boldsymbol{x}) \simeq H(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \tag{2.4}$$

By minimizing W_k, we can find a point where the gradient $\boldsymbol{g}$ becomes small. To minimize W_k, we solve:

$$\begin{bmatrix} B_k & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{c} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \quad \boldsymbol{c} := \begin{bmatrix} c_k \\ c_{k-1} \\ \vdots \\ c_{k-L+1} \end{bmatrix}, \quad \mathbf{1} := \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad B_k := \begin{bmatrix} \boldsymbol{e}_k^\top \boldsymbol{e}_k & \dots & \boldsymbol{e}_k^\top \boldsymbol{e}_{k-L+1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{e}_{k-L+1}^\top \boldsymbol{e}_k & \dots & \boldsymbol{e}_{k-L+1}^\top \boldsymbol{e}_{k-L+1} \end{bmatrix} \tag{2.5}$$

Although the RMM-DIIS method improves convergence behavior, it tends to converge to saddle points when used directly for function optimization, as it searches for points where the gradient becomes zero. Given that high-symmetry phases can be unstable saddle points at low temperatures in structural phase transition calculations, there is a risk of convergence to structures that are not the most stable at each temperature, which is undesirable.

2.3 BFGS Method

Similarly, the BFGS (Broyden–Fletcher–Goldfarb–Shanno) method is an effective approach when the Hessian cannot be calculated directly [5]. The optimization update is performed as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H_k^{-1}\boldsymbol{g}(\boldsymbol{x}_k) \tag{2.6}$$

$$H_{k+1}^{-1} = H_k^{-1} + \frac{\boldsymbol{y}_k^\top \boldsymbol{s}_k + \boldsymbol{y}_k^\top H_k^{-1}\boldsymbol{y}_k}{\left(\boldsymbol{y}_k^\top \boldsymbol{s}_k\right)^2}\boldsymbol{s}_k\boldsymbol{s}_k^\top - \frac{H_k^{-1}\boldsymbol{y}_k\boldsymbol{s}_k^\top + \boldsymbol{s}_k\boldsymbol{y}_k^\top H_k^{-1}}{\boldsymbol{y}_k^\top \boldsymbol{s}_k} \tag{2.7}$$

$$\boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \quad \boldsymbol{y}_k := \boldsymbol{g}(\boldsymbol{x}_{k+1}) - \boldsymbol{g}(\boldsymbol{x}_k) \tag{2.8}$$

However, updates are only performed at points that satisfy the condition $\boldsymbol{y}_k^\top \boldsymbol{s}_k > 0$, which ensures the positive definiteness of the Hessian and prevents convergence to saddle points.

2.4 GDIIS Method

The focus of this study is on the GDIIS (Geometry optimization using Direct Inversion in the Iterative Subspace) method, which combines the RMM-DIIS and BFGS methods [6]. In this study, we used the result of BFGS as an approximation for the residual vector in RMM-DIIS:

$$\boldsymbol{g}(\boldsymbol{x}_k) \to H_k^{-1}\boldsymbol{g}(\boldsymbol{x}_k) \tag{2.9}$$

Additionally, a condition was imposed on the direction of the new search point obtained by the RMM-DIIS procedure. Specifically, the update is allowed only if the angle θ_k, defined as:

$$\cos\theta_k = \frac{\boldsymbol{\Delta}_{\mathrm{BFGS}}^\top \boldsymbol{\Delta}_{\mathrm{DIIS}}}{|\boldsymbol{\Delta}_{\mathrm{BFGS}}||\boldsymbol{\Delta}_{\mathrm{DIIS}}|}, \quad \boldsymbol{\Delta}_{\mathrm{BFGS}} := -H_k^{-1}\boldsymbol{g}(\boldsymbol{x}_k), \quad \boldsymbol{\Delta}_{\mathrm{DIIS}} := \boldsymbol{x}_{k+1} - \boldsymbol{x}_i \tag{2.10}$$

is sufficiently close. By imposing this condition, the method suppresses convergence to saddle points. Specifically, we adopted the following thresholds from previous research [6]:

$$\cos\theta > \mathrm{Threshold}(L) := \begin{cases} 0.97 & (L = 2) \\ 0.84 & (L = 3) \\ 0.71 & (L = 4) \\ 0.67 & (L = 5) \\ 0.62 & (L = 6) \\ 0.56 & (L = 7) \\ 0.49 & (L = 8) \\ 0.41 & (L = 9) \\ 0.00 & (L \geq 10) \end{cases} \tag{2.11}$$

where L is the number of past points used in RMM-DIIS scheme.

3 Results and Discussion

3.1 Test Case

In this section, we verify the optimization method described in the previous chapter. This study intends to apply the developed methods to structure optimization at finite temperatures, particularly for calculating structural phase transitions. In structural phase transitions, it is typical for the transition to occur between a highly symmetric high-temperature phase and a low-symmetry low-temperature phase. When plotting the free energy landscape along the soft mode, as shown in Figure 1, the high-symmetry phase becomes stable at high temperatures ($T > T_c$), while at low temperatures ($T < T_c$), the high-symmetry phase becomes an unstable saddle point, leading to a phase transition to the low-symmetry phase. Therefore, we will verify the method using a test case involving a double-well function that includes a saddle point, as shown below:

$$f(\boldsymbol{x}) = \exp\big((3x^4 + 4x^3 - 12x^2 + 32)/32\big)(1.0 + y^2) \tag{3.1}$$

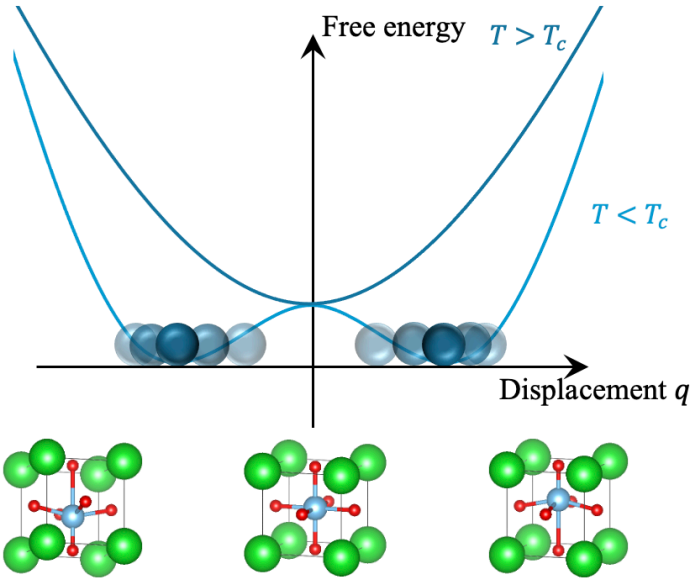

Figure 1 Schematic diagram of the temperature dependence of the free energy landscape along the soft mode during a structural phase transition.

3.2 Results of RMM-DIIS

First, the optimization problem for the test case was solved using RMM-DIIS, and the results are shown in Figure 2. In Figure 2, the solution converged to the saddle point of the double well, and although it converged to a point where the gradient was zero, it did not meet the objective of finding a local minimum. This corresponds to the unstable high-symmetry phase at $T < T_c$ in the previous example, which is undesirable behavior for the calculation of structural phase transitions targeted in this study.

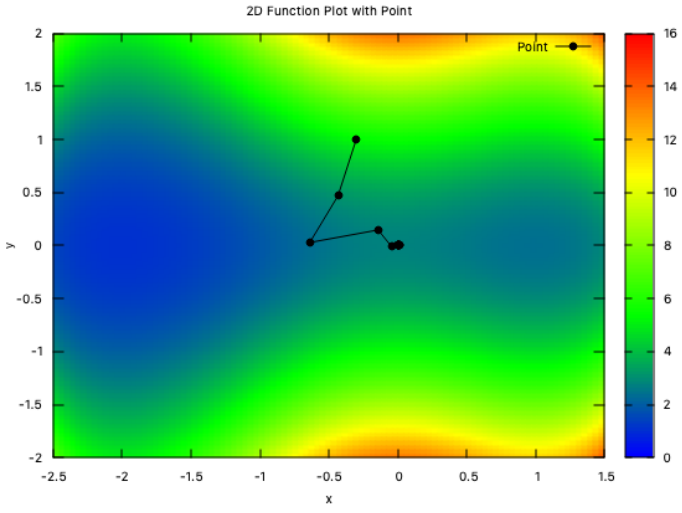

Figure 2 RMM-DIIS ($L = 6$) failing to calculate the minimum for the double-well potential. The function converged to the saddle point located at $\boldsymbol{x}_{\text{saddle}} = (0, 0)$. The initial values used for the calculation were $(-0.3, 1.0)$.

3.3 Results of GDIIS

Next, the performance of GDIIS, a method that can potentially address the issue in the previous section, was examined.

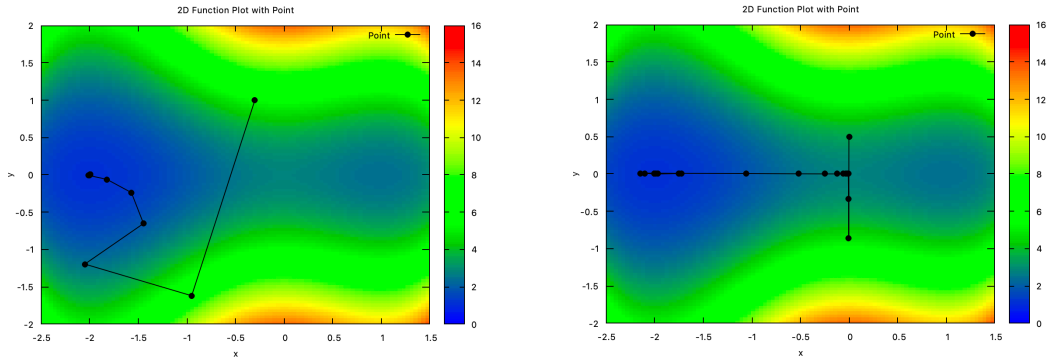

Figure 3 Optimization using GDIIS ($L = 6$). The initial values were $(-0.3, 1.0)$ for the left figure and $(-0.001, 0.5)$ for the right figure. Even in the same situation where RMM-DIIS failed, the function can be seen updating toward smaller values. In the actual calculation, the initial Hessian matrix was set as $H_0^{-1} = 0.5I$.

When the optimization was performed with GDIIS under the same conditions where RMM-DIIS converged to the saddle point, GDIIS successfully avoided converging to the saddle point and headed toward the minimum point, as seen in Figure 3. This is because the condition that restricts the direction of DIIS changes the step in the direction of the saddle point to a standard BFGS step. Furthermore, this effect is significant, as GDIIS exhibited correct behavior even with initial conditions close to the saddle point, as shown in Figure 3.

Next, we want to compare the performance difference between GDIIS and the BFGS methods, which could also address the issue from the previous section. GDIIS is considered to converge more quickly than BFGS in flat regions of functions far from extrema. Here, as an example of a flat function, we benchmark the function $f(\boldsymbol{x}) = \boldsymbol{x}^4$. Since many optimization methods assume that the optimal value is quadratic, optimization is more challenging for systems where the minimum is expressed by a cubic or higher order.

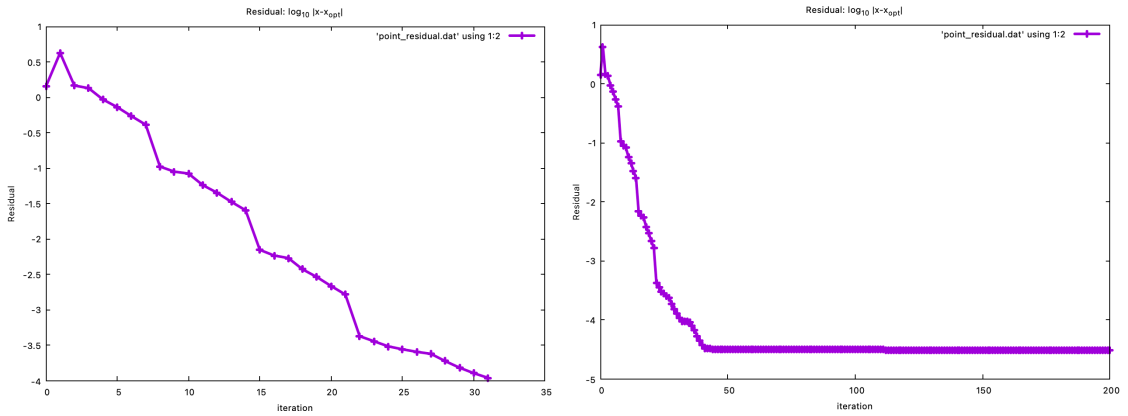

Figure 4 Optimization of $f(\boldsymbol{x}) = \boldsymbol{x}^4$ using GDIIS. The vertical axis shows the distance from the exact minimum $\boldsymbol{x} = (0, 0)$ plotted on a log scale. The left figure shows the graph until the error reaches 10^{-4}, and the right figure shows the results after 200 iterations.

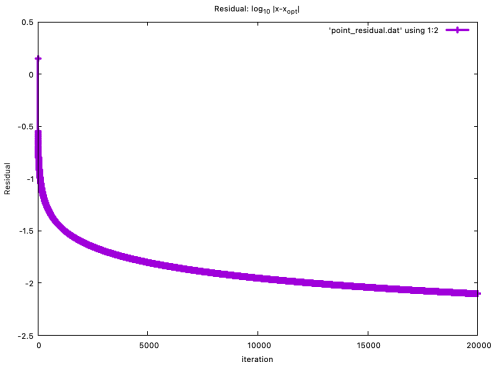

Figure 5 Exploration of $f(\boldsymbol{x}) = \boldsymbol{x}^4$ using the gradient descent method.

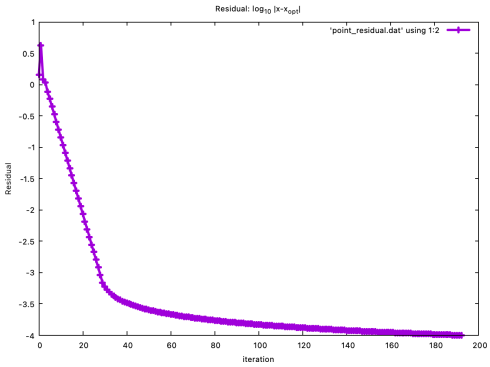

Figure 6 Exploration of $f(\boldsymbol{x}) = \boldsymbol{x}^4$ using the BFGS method.

Figures 4, 5, and 6 show the number of steps taken to reach an error of approximately 10^{-4} from the optimal value. Notably, GDIIS approached the desired error in approximately 30 iterations, while BFGS took about 190 iterations, and gradient descent failed to reach 10^{-2} after 20,000 iterations. This confirms that GDIIS is more suitable for optimizing such flat functions compared to other optimization methods.

4 Implementation into the ALAMODE Package

To utilize the methods validated in this research within the ALAMODE package, modifications to the ALAMODE code are necessary. Currently, the calculation of anharmonic phonon theory and the gradient of free energy is invoked as a single function during each step of structural optimization, updating the structure accordingly. Moreover, since each method is implemented in a way specific to this problem, a deep understanding of the details of the ALAMODE code is required for algorithm implementation, making it challenging to easily test various methods.

Therefore, in this research, we focused on decoupling the part of the code that solves general optimization problems, enabling a more straightforward switching between different optimization engines for testing purposes. We began preparations for test cases to ensure that the modified code worked correctly, but we were unable to complete the implementation within the given timeframe. Applying the validated algorithms to actual materials remains a future challenge.

5 Conclusion and Future Outlook

In this research, we confirmed that using the GDIIS method allows for robust convergence to local minima and efficient convergence in flat functions simultaneously. Regarding the implementation into the ALAMODE package, we worked on modifying the implementation to allow for switching between optimization algorithms, but we were unable to test the validated algorithms on actual materials.

We aim to incorporate GDIIS into the ALAMODE code and conduct benchmarks on several materials, starting with $BaTiO_3$, to provide this as a new feature. Additionally, while this method adopted BFGS and RMM-DIIS, the original paper on GDIIS [6] mentions methods with higher stability. We would like to continue working on the validation of these algorithms.

6 Acknowledgments

In this research, We would like to express our sincere gratitude to our supervising professors, Prof. Shinji Tsuneyuki and Prof. Ryotaro Arita from the Department of Physics at the University of Tokyo and Dr. Terumasa Tadano from the National Institute for Materials Science, for their tremendous support and guidance. Additionally, I would like to thank Prof. Taisuke Ozaki from

the Institute of Solid State Physics at the University of Tokyo, who served as one of our co-supervisors, for approving this collaborative research, as well as the MERIT program for providing the opportunity for this fusion research.

References

[1] Ryota Masuki, Takuya Nomoto, Ryotaro Arita, and Terumasa Tadano. Ab initio structural optimization at finite temperatures based on anharmonic phonon theory: Application to the structural phase transitions of batio 3. Physical Review B, 106(22):224104, 2022.

[2] Ryota Masuki, Takuya Nomoto, Ryotaro Arita, and Terumasa Tadano. Anharmonic grüneisen theory based on self-consistent phonon theory: Impact of phonon-phonon interactions neglected in the quasiharmonic theory. Physical Review B, 105(6):064112, 2022.

[3] Ryota Masuki, Takuya Nomoto, Ryotaro Arita, and Terumasa Tadano. Full optimization of quasiharmonic free energy with an anharmonic lattice model: Application to thermal expansion and pyroelectricity of wurtzite gan and zno. Physical Review B, 107(13):134119, 2023.

[4] Peter Pulay. Improved scf convergence acceleration. Journal of Computational Chemistry, 3(4):556–560, 1982.

[5] Charles George Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. IMA Journal of Applied Mathematics, 6(1):76–90, 1970.

[6] Ödön Farkas and H Bernhard Schlegel. Methods for optimizing large molecules. part iii. an improved algorithm for geometry optimization using direct inversion in the iterative subspace (gdiis). Physical Chemistry Chemical Physics, 4(1):11–15, 2002.

[7] Ödön Farkas and H Bernhard Schlegel. Methods for geometry optimization of large molecules. i. an o(n^2) algorithm for solving systems of linear equations for the transformation of coordinates and forces. The Journal of chemical physics, 109(17):7100–7104, 1998.

[8] Ödön Farkas and H Bernhard Schlegel. Methods for optimizing large molecules. ii. quadratic search. The Journal of Chemical Physics, 111(24):10806–10814, 1999.