Journal of Information Processing Vol.23 No.3 366–372 (May 2015)

[DOI: 10.2197/ipsjjip.23.366]

Regular Paper

Improvement of Hierarchical Matrices with Adaptive Cross Approximation for Large-scale Simulation

Akihiro Ida[1,2,a)] Takeshi Iwashita[2,3,b)] Makiko Ohtani[4,c)] Kazuro Hirahara[2,4,d)]

Received: July 18, 2014, Accepted: November 21, 2014

Abstract: We propose an improved method for hierarchical-matrices (H-matrices) using adaptive cross approximation (ACA) as the low-rank approximation. The improvement consists of a kind of normalization and a new stopping criterion for the ACA. By using the proposed method, we can avoid the trouble that ranks of approximated matrices increase rapidly as the matrix size increases when the conventional H-matrices with ACA are employed to an integral equation whose kernel function has high-order singularities. In particular, application of the proposed method enables us to perform large-scale simulations such that the conventional H-matrices with ACA fail to construct the low-rank approximation. Applicability of the proposed method is confirmed through numerical experiments on an earthquake cycle simulation.

Keywords: low-rank approximation, hierarchical matrices, adaptive cross approximation, integral equation method, earthquake cycle simulation

1. Introduction

The integral equation method represented by the boundary-element method (BEM) is one of the important numerical methods applied widely to scientific analyses. Naïve application of the BEM yields linear equation systems with dense coefficient matrices. This requires a memory footprint proportional to n^2 and a computational effort of $O(n^2)$ or $O(n^3)$, where n is the number of unknowns. The same issues also arise in the n-body problem. In recent years, large-scale BEM analyses have been conducted thanks to advanced computer technology or the use of approximation techniques for dense coefficient matrices.

A number of approximation techniques for the above problem have been proposed including hierarchical-matrices (H-matrices) [1], [2], [3], fast multi-pole [4], [5], [6], and tree [7] methods. All these methods assume that the kernel function of the integral operator has singularities of the form $g(x, y) = |x - y|^{-p}$,where $p > 0$, or $g(x, y) = \log|x - y|$, and are based on the same idea that the kernel function for remote two points x, y can be approximated by a degenerate kernel. That is, the kernel function $g(x, y)$ is possibly expanded as $g(x, y) \cong \sum_{\nu=1}^{k} g_1^{\nu}(x) g_2^{\nu}(y)$ for a positive integer k, so that integration with respect to the x-variable can be separated from that with respect to the y-variable. Most of these techniques require knowledge of the concrete form of the kernel function, and this mathematical computation must be done by hand. It is sometimes difficult or very complicated to obtain the form of such a degenerate kernel. For these problems, skeleton approximation [8] or adaptive cross approximation (ACA) [9], [10] are available. These methods perform an algebraic approximation corresponding to the quadrature of the degenerate kernel; i.e., only the original entries in the coefficient matrix are used in this approximation. ACA and H-matrices have an affinity with each other. In H-matrices, the approximation mentioned above is expressed as low-rank sub-matrices. By using a cluster tree with some geometrical criterion, H-matrices find a suitable permutation and partition such that numerous rather large sub-matrices become low-rank matrices. **Figure 1**, transcribed from Ref. [11], shows visualization examples of H-matrices. ACA is utilized to approximate each low-rank sub-matrix.

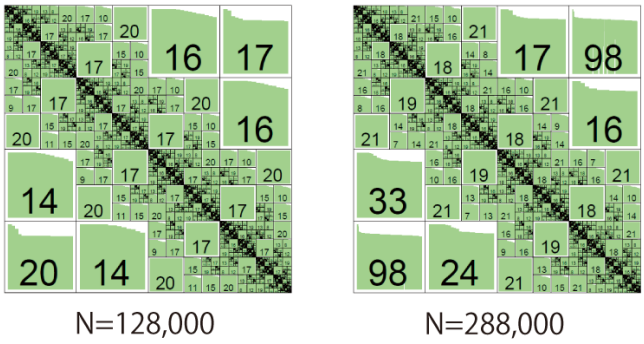


Fig. 1 H-matrices derived from a quasi-dynamic equilibrium equation in the earthquake cycle simulation. Each cell shows a sub-matrix, with the number representing the rank of the approximated sub-matrix [11].

H-matrices with ACA have been proven to be very effective when applied to practical applications [9], [11], [12], [13]. However, as pointed out in Ref. [14], the stopping criterion generally

[1] ACCMS, Kyoto University, Kyoto 606–8501, Japan
[2] JST CREST, Kawaguchi, Saitama 332–0012, Japan
[3] Information Initiative Center, Hokkaido University, Sapporo, Hokkaido 060–0811, Japan
[4] Graduate School of Science, Kyoto University, Kyoto 606–8501, Japan
[a)] ida@media.kyoto-u.ac.jp
[b)] iwashita@iic.hokudai.ac.jp
[c)] otani.makiko.63n@st.kyoto-u.ac.jp
[d)] hirahara@kugi.kyoto-u.ac.jp

used in ACA is not efficient; that is, it is unnecessarily strict when the kernel function has higher-order singularities. Furthermore, as reported by Ohtani et al. [11], the computational time and memory usage increase rapidly as the matrix size n increases when an H-matrix with ACA is applied to an earthquake cycle simulation, owing to the fact that the ranks of the far-off-diagonal sub-matrices are increased. In the example shown in Fig. 1, we can see that rank 17 of the upper-outermost sub-matrix for $n = 128{,}000$ jumps to 98 for $n = 288{,}000$. It is expected that in the case of a much larger n, ACA would fail to construct efficient approximations. This is a bottleneck for very large-sized simulations. To remedy this inconvenience, Ohtani et al. [11] introduced an upper-limit on the ranks of approximated sub-matrices by ACA, and searched for a sufficient range of the upper-limit. They successfully completed the simulation without losing simulation accuracy with the upper-limit set to 20, despite the ACA not satisfying the stopping criterion.a However, it is generally difficult to determine the upper-limit accurately because it depends on various conditions such as the matrix size n, structure of the H-matrix, boundary elements, and physical parameters.

In this paper, we discuss an improvement for H-matrices using ACA to avoid the inconvenience mentioned above. In Section 2, we formulate the linear equation system derived from integral equations with a singular kernel and apply H-matrices to the coefficient matrix of the linear equation system. In Section 3, we discuss the controversial points of conventional H-matrices with ACA, and propose an improved method for it. Numerical experiments of an earthquake cycle simulation are presented in Section 4. The last section is devoted to the conclusion.

2. Linear Equation System Derived from Integral Equations with Singular Kernels

Let H be a Hilbert space of functions on a $(d-1)$-dimensional domain $\Omega \subset \mathbb{R}^d$ and H' the dual space of H. For $u \in H, f \in H'$ and a kernel function of a convolution operator $g : \mathbb{R}^d \times \Omega \to \mathbb{R}$, we consider the integral equation

$$\int_\Omega g(x,y)u(y)\mathrm{d}y = f. \tag{1}$$

To calculate Eq. (1) numerically, we divide domain Ω into elements $\Omega^h = \{\omega_j : j \in J\}$, where J is an index set. When using weighted residual methods such as the Ritz-Galerkin method and the collocation method, function u is approximated from an n-dimensional subspace $H^h \subset H$. Given a basis $(\varphi_i)_{i\in\beth}$ of H^h for an index set $\beth := \{1, \cdots, n\}$, the approximant $u^h \in H^h$ to u can be written using a coefficient vector $\phi = (\phi_i)_{i\in\beth}$ satisfying $u^h = \sum_{i\in\beth} \phi_i\varphi_i$. We note that supports of the basis $\Omega^h_{\varphi_i}$:=supp φ_i are assembled from sets ω_j. Equation (1) is reduced to the following linear equation system:

$$a\phi = b. \tag{2}$$

In the case of the Ritz-Galerkin method, entries of a and b are given by

$$a_{ij} = \int_\Omega \varphi_i(x) \int_\Omega g(x,y)\varphi_j(y)\mathrm{d}y\mathrm{d}x \quad \text{for all } i,j \in \beth, \tag{3}$$

$$b_i = \int_\Omega \varphi_i(x) f \mathrm{d}x \qquad \text{for all } i \in \beth. \tag{4}$$

When the kernel function g on the left-hand side of Eq. (1) has a local support, coefficient matrix $a \in \mathbb{R}^{\beth\times\beth}$ will be sparse. However, we have to deal with dense matrices for coefficient matrix a when the kernel functions are written in the form $g(x,y) = |x-y|^{-p}$,where $p > 0$. Support for this type of kernel function is the whole of domain Ω. Such kernel functions appear in a number of scientific applications, for example, electric field analyses, mechanical analyses, and earthquake cycle simulation.

Though coefficient matrix a has no explicit structure such as a band or block structure, we can find implicit structures under the following assumption. Hereafter, we assume that kernel function $g(x,y)$ is a singular kernel satisfying

$$g(x,y) \in \mathrm{span}(\{|x-y|^{-p}, p > 0\}), \tag{5}$$

where span$(\cdot)$ means the set of all finite linear combinations. We now suppose two subsets (clusters) $s, t \subset \beth$ and their corresponding domains defined as

$$\Omega^h_s := \bigcup_{i\in s} \mathrm{supp}\ \varphi_i, \quad \Omega^h_t := \bigcup_{i\in t} \mathrm{supp}\ \varphi_i. \tag{6}$$

We refer to the cluster pair (s,t) as 'admissible', if

$$\min\{\mathrm{diam}(\Omega^h_s),\ \mathrm{diam}(\Omega^h_t)\} \le \eta \mathrm{dist}(\Omega^h_s, \Omega^h_t), \tag{7}$$

where diam$(\cdot)$ is the Euclidean diameter of a set and dist$(\cdot)$ denotes the Euclidean distance between two sets and $\eta > 0$ is a parameter determined by the singularities of the problem. On a domain corresponding to admissible cluster pairs, the kernel function can be approximated with a certain degree of accuracy by a degenerate kernel such as $g(x,y) \cong \sum_{\nu=1}^{k} g_1^\nu(x) g_2^\nu(y)$, where k is a positive number. Then the corresponding sub-matrix $a|_{s\times t}$ has a low-rank if we construct $a|_{s\times t}$ by gathering those columns and rows in coefficient matrix a with ordinal numbers belonging to sets s and t. Under our assumptions, it is possible to find a permutation and a partition of index set $\beth$ such that there are numerous low-rank sub-matrices with fairly large sizes. These manipulations, the permutation, and the partition, are called 'clustering'. It is known that clustering is efficiently performed by constructing a cluster tree,$T_\beth$, based on the distance between the clusters and the size of the corresponding supports. The cluster tree induces the block cluster tree, $T_{\beth\times\beth}$, whose leaves $\mathcal{L}(T_{\beth\times\beth})$ form a partition of $\beth \times \beth$. An H-matrix is defined using the block cluster tree $T_{\beth\times\beth}$ [3].

Next we consider approximating coefficient matrix a by the leaves of H-matrices with a low-rank approximation method. Hereafter, we assume that proper clustering of index set $\beth$, which is induced by cluster tree $T_\beth$ and block cluster tree $T_{\beth\times\beth}$, is performed; that is, coefficient matrix $a \in \mathbb{R}^{\beth\times\beth}$ has a structure comprising sub-matrices as shown in Fig. 1. The sub-matrices $a|_{s\times t}, (s,t) \in \mathcal{L}(T_{\beth\times\beth})$ are divided into two groups based on whether the corresponding cluster pair is admissible. We write the approximated coefficient matrix $\tilde{a} \in \mathbb{R}^{\beth\times\beth}$ as

$$\tilde{a}|_{s\times t} = \begin{cases} \tilde{a}^{s,t}_{k_{s,t}} & \text{if } (s,t) \text{ is admissible,} \\ a|_{s\times t} & \text{otherwise,} \end{cases} \quad \text{for all } (s,t) \in \mathcal{L}(T_{\beth\times\beth}), \tag{8}$$

where $\tilde{a}^{s,t}_{k_{s,t}} \in \mathbb{R}^{s\times t}$ represents the approximated sub-matrix of $a|_{s\times t}$ with rank $k_{s,t}$; i.e.,

$$\tilde{a}^{s,t}_{k_{s,t}} = VW^T = \sum_{\nu=1}^{k_{s,t}} v^\nu (w^\nu)^T, \tag{9}$$

where $V \in \mathbb{R}^{s\times k_{s,t}}$, $W \in \mathbb{R}^{t\times k_{s,t}}$, $v^\nu \in \mathbb{R}^{|s|}$ and $w^\nu \in \mathbb{R}^{|t|}$. Then, the number of entries $N_{\tilde{a}}$ in the approximated coefficient matrix $\tilde{a}$ is described by using the sizes of sub-matrices and ranks thereof :

$$N_{\tilde{a}} = \sum_{(s,t)\in\mathcal{L}(T_{\mathcal{I}\times\mathcal{I}})} N_{\tilde{a}|_{s\times t}},$$
$$\text{where } N_{\tilde{a}|_{s\times t}} = \begin{cases} k_{s,t}(\#s + \#t) & \text{if } (s,t) \text{ is admissible,} \\ \#s\#t & \text{otherwise.} \end{cases} \tag{10}$$

The computational effort for arithmetic in the use of H-matrices, such as a construction of an H-matrix and a multiplication of an H-matrix and a vector, depends on the number of entries $N_{\tilde{a}}$.

The singular value decomposition might give the approximated sub-matrix $\tilde{a}^{s,t}_{k_{s,t}}$ in the form of Eq. (9) with the lowest rank. However, it is too expensive in terms of computational cost and requires all the entries of $\tilde{a}|_{s\times t}$. As an efficient alternative, Goreinov et al. [8] proposed the method which only requires pivot columns and pivot rows of $\tilde{a}|_{s\times t}$. They showed the following existence theorem for $\tilde{a}^{s,t}_{k_{s,t}}$.

Theorem 1 *Let $a|_{s\times t}, R \in \mathbb{R}^{s\times t}$ be matrices with $\|a|_{s\times t} - R\| \le \hat{\varepsilon}$ and rank(R) $\le \hat{k}$. Then there exist a subset $s^* \subset s$ of pivot rows, a subset $t^* \subset t$ of columns, and a matrix $S \in \mathbb{R}^{s^*\times t^*}$ with*

$$\|a|_{s\times t} - \tilde{a}^{s,t}_{k_{s,t}}\|_2 \le \hat{\varepsilon}(1 + 2\sqrt{k}(\sqrt{\#s} + \sqrt{\#t})),\ \tilde{a}^{s,t}_{k_{s,t}} = a|_{s\times t^*} \cdot S \cdot a|_{s^*\times t}.$$

All the methods based on Theorem 1 are collectively called the 'Cross Approximation'. There are some variants of the cross approximation, such as ACA and ACA+. These variants differ in the strategy for selecting the sets of pivot rows s^* and pivot columns t^* and the stopping criterion of the algorithm. In most of ACA and its variants, a pivot column and a pivot row is alternately selected one vector by one vector until the stopping criterion is satisfied. Then, $\#s^* = \#t^* = k_{s,t}$ and the matrix S in Theorem1 can be written as $S = a|_{t^*\times s^*}{}^{-1}$, i.e., $\tilde{a}^{s,t}_{k_{s,t}} = UW^T = a|_{s\times t^*} \cdot a|_{t^*\times s^*}{}^{-1} \cdot a|_{s^*\times t}$. On the other hand, the inequality in Theorem 1 will be used in the next section to evaluate the error caused by the approximation for $a|_{s\times t}$ by $\tilde{a}^{s,t}_{k_{s,t}}$.

Since the above discussion on coefficient matrix a depends on the kernel function and localization of the base functions, these hold if we use weighted residual methods other than the Ritz-Galerkin method.

3. Proposed Method

The aim of this section is to propose a method for calculating Eq. (2) with the approximated coefficient matrix described by Eq. (8) without the trouble mentioned in Section 1. We first discuss the controversial points of conventional H-matrices with ACA. Our ideas for improvement of the conventional method are explained along with the discussion. In the end, we propose the improved method.

As mentioned in Section 1, the rank of the approximated matrix $k_{s,t}$ increases rapidly as the size of the problem n increases, when conventional H-matrices with ACA are applied to the earthquake cycle simulation [11] with a third order singular kernel. We contrived the method to prevent rank $k_{s,t}$ from increasing rapidly, under suppositions that such a problem could arise for the following two reasons:

(i) The stopping criterion of ACA is uniform over the whole matrix.

(ii) There is a large difference in the absolute value of entries between rows.

First, we consider (i). Most numerical simulations require that the approximated matrix $\tilde{a}$ satisfies

$$\frac{\|a - \tilde{a}\|_F}{\|a\|_F} \le \varepsilon \ \text{ or } \ \frac{\|a - \tilde{a}\|_2}{\|a\|_2} \le \varepsilon, \tag{11}$$

where $\varepsilon \in \mathbb{R}_{>0}$ is the given error tolerance and $\|\cdot\|_F$ denotes the Frobenius norm. Furthermore, in the case that the sub-matrices $a|_{s\times t}$ are approximated by ACA, the following inequality is conventionally used as the stopping criterion of the algorithm,

$$\frac{\|a|_{s\times t} - \tilde{a}|_{s\times t}\|_F}{\|a|_{s\times t}\|_F} \le \varepsilon. \tag{12}$$

We can easily prove that Eq. (11) is satisfied if Eq. (12) holds. Criterion Eq. (12) is uniform over the whole matrix; i.e., it does not depend on the location of the sub-matrix. The denominator on the left-hand side of Eq. (12), $\|a|_{s\times t}\|_F$, is expected to be much smaller for far-off-diagonal sub-matrices than for on-diagonal ones of matrices arising from the singular kernel Eq. (5), especially one with high-order singularities. Criterion Eq. (12) requires a greater accuracy for far-off-diagonal sub-matrices than for on-diagonal ones. The approximation errors do not arise from on-diagonal or near-diagonal sub-matrices considering Eq. (8). For the whole matrix comprising approximated sub-matrices with criterion Eq. (12), the left-hand side of Eq. (11) is possibly much smaller than tolerance ε. Moreover, since it is rare that $\|a|_{s\times t}\|_F$ is calculated exactly in Eq. (12), $\|\tilde{a}|_{s\times t}\|_F$ is used instead in practical applications. According to the expression in Theorem 1, this replacement causes the error, which is including the square root of the sub-matrix size.

Next, we discuss condition (ii). We suppose this situation occurs if the element sizes are not uniform. Let $\tilde{g}(x,y)$ be an approximation function of kernel function $g(x,y)$ in Eq. (1). For a given tolerance $\tilde{\varepsilon} \in \mathbb{R}_{>0}$ and any $\sigma, \tau \subseteq \Omega$, we assume that $\tilde{g}(x,y)$ satisfies

$$|g(x,y) - \tilde{g}(x,y)| \le \tilde{\varepsilon} \quad \text{for all } x \in \sigma, y \in \tau. \tag{13}$$

Börm et al. [3] presented the following two theorems given this assumption.

Theorem 2 *Let $\mu_{\max} \in \mathbb{R}_{>0}$ be a constant satisfying*

$$\|u\|^2_{H^h} \le \mu_{\max}\|u\|^2_2 \quad \textit{for all } t \in T_{\mathcal{I}} \textit{ and } u \in \mathbb{R}^t.$$

Moreover, let $s, t \in T_{\mathcal{I}}$ be clusters satisfying $\Omega^h_s \subseteq \sigma$ and $\Omega^h_t \subseteq \tau$. Then we have

$$\|a|_{s\times t} - \tilde{a}^{s,t}\|_2 \le \mu_{\max}\tilde{\varepsilon}\sqrt{|\Omega^h_s||\Omega^h_t|}. \tag{14}$$

Theorem 3 *Let $C_{\text{ov}} \in \mathbb{N}$ be a constant satisfying*

$$\#\{i \in \sqsupset : \varphi_i(x) \neq 0\} \leq C_{\mathrm{ov}} \ \textit{for all } x \in \Omega^h.$$

For matrix a and its approximation $\tilde{a}$ satisfying Eq. (8), we have

$$\|a - \tilde{a}\|_2 \leq C_{\mathrm{ov}} \mu_{\max} |\Omega^h| \tilde{\varepsilon}, \tag{15}$$

where $|\cdot|$ denotes the support size.

Börm et al. [3] remarked that the optimal choice for constant C_{ov} is given by the number of base functions associated with the element. For piecewise constant functions we have $C_{\mathrm{ov}} = 1$, whereas piecewise linear basis functions on triangles in $\mathbb{R}^3$ lead to the choice $C_{\mathrm{ov}} = 3$. In the proof of Theorem 3, the inequality $\sum_{(s,t)\in\mathcal{L}(T_{\sqsupset\times\sqsupset})} \sqrt{|\Omega_s^h||\Omega_t^h|} \leq C_{\mathrm{ov}}|\Omega^h|$ is used in addition to Theorem 2. We derive the error upper bound of the whole coefficient matrix approximation from the assumption for the kernel function via the error upper bound of the sub-matrix approximation. Inspired by these theorems, we here introduce an inequality:

$$\|a|_{s\times t} - \tilde{a}_{k_{s,t}}^{s,t}\| \leq \frac{\sqrt{|\Omega_s^h||\Omega_t^h|}}{C_{\mathrm{ov}}|\Omega^h|}\varepsilon. \tag{16}$$

When Eq. (16) holds, $\|a - \tilde{a}\|_2 \leq \varepsilon$ is satisfied as follows:

$$\begin{aligned}\|a - \tilde{a}\|_2 &\leq \sqrt{\sum_{(s,t)\in\mathcal{L}(T_{\sqsupset\times\sqsupset})} \|a|_{s\times t} - \tilde{a}_{k_{s,t}}^{s,t}\|_2^2} \\ &\leq \frac{\varepsilon}{C_{\mathrm{ov}}|\Omega^h|} \sum_{(s,t)\in\mathcal{L}(T_{\sqsupset\times\sqsupset})} \sqrt{|\Omega_s^h||\Omega_t^h|} \\ &\leq \varepsilon. \end{aligned} \tag{17}$$

Inequality Eq. (16) implies a relation between the element sizes and approximation accuracy of the sub-matrix. On the other hand, element sizes do not appear in criterion Eq. (12). Criterion Eq. (12) requires a greater accuracy for sub-matrices corresponding to basis supports that include large sized elements than for ones with small sized elements.

Criterion Eq. (12) is more suitable than Eq. (12) for the problem formulated in Section 2. Our idea is the use of inequality Eq. (16) for the stopping criterion of ACA instead of the conventional criterion Eq. (12). However, it is inconvenient to directly apply the criterion Eq. (16) to ACA in practical applications for the following three reasons.

(I) We have to handle element sizes corresponding to the basis supports.

(II) It is expensive to compute the 2-norm, but not the Frobenius norm.

(III) We may have to investigate physical quantities to determine tolerance ε because the criterion Eq. (16) is not given in the form of a relative error.

To overcome the above three problems, we here try to simulate the criterion Eq. (16) by normalization and the use of sizes of (sub-)matrices easy to handle more than element sizes. We here introduce the normalization of Eq. (2) using the diagonal and the size of the coefficient matrix as follows:

$$A\Phi = B, \tag{18}$$

$$\begin{aligned}&A = D_n^{-T} a D_n^{-1}, \Phi = D_n\phi \text{ and } B = D_n^{-T} b \\ &\text{for } D_n = \mathrm{diag}\left(\sqrt{\tfrac{a_{11}}{\sqrt{n}}}, \sqrt{\tfrac{a_{22}}{\sqrt{n}}}, \cdots, \sqrt{\tfrac{a_{nn}}{\sqrt{n}}}\right).\end{aligned} \tag{19}$$

The problem (I) is dealt with by the normalization. In many practical simulations, the absolute values of diagonal entries a_{ii} are proportional to the element sizes if the kernel function $g(x, y)$ is a function of distance $|x-y|$ only. Then, the support sizes $|\Omega_s^h|$, $|\Omega_t^h|$, and $|\Omega^h|$ can be replaced by the number of bases in clusters $\#s$, $\#t$, and n, respectively, after the normalization. The second problem is addressed by simply replacing the kind of norm. This replacement does not change the criterion much. Its influence is slight considering formula $\|\cdot\|_2 \leq \|\cdot\|_F \leq \sqrt{k}\|\cdot\|_2$ because rank k is supposed to be a small number [15]. Then, the criterion Eq. (16) corresponding to the normalized equation can be written as

$$\|A|_{s\times t} - \tilde{A}_{k_{s,t}}^{s,t}\|_F \leq \frac{\sqrt{\#s\#t}}{C_{\mathrm{ov}} n}\varepsilon. \tag{20}$$

Finally, we confirm that the problem (III) is also already dealt with. It is shown that criterion Eq. (11) is satisfied.

$$\begin{aligned}\|A - \tilde{A}\|_F^2 &\leq \sum_{(s,t)\in\mathcal{L}(T_{\sqsupset\times\sqsupset})} \|A|_{s\times t} - \tilde{A}_{k_{s,t}}^{s,t}\|_F^2 \\ &\leq \varepsilon^2 \frac{\sum_{(s,t)\in\mathcal{L}(T_{\sqsupset\times\sqsupset})} \#s\#t}{C_{\mathrm{ov}}^2 n^2} \\ &\leq \varepsilon^2 = \varepsilon^2\|D\|_F^2 \leq \varepsilon^2\|A\|_F^2 \\ \Rightarrow \frac{\|A - \tilde{A}\|_F}{\|A\|_F} &\leq \varepsilon. \end{aligned} \tag{21}$$

Note that $D = \mathrm{diag}(A)$ and $\|D\|_F = 1$. For more convenient use of the criterion Eq. (19) in ACA, we rewrite the left hand side of Eq. (20):

$$\|v^{k_{s,t}}\|_2 \|w^{k_{s,t}}\|_2 \leq \frac{\sqrt{\#s\#t}}{C_{\mathrm{ov}} n}\varepsilon, \tag{22}$$

where $v^{k_{s,t}}$ and $w^{k_{s,t}}$ are the latest vectors of sequences v^ν, w^ν, respectively, generated in the ACA as $\tilde{A}_{k_{s,t}}^{s,t} = \sum_{\nu=1}^{k_{s,t}} v^\nu (w^\nu)^T$. This displacement depends on the following heuristic used usually in ACA.

$$\begin{aligned}\|A|_{s\times t} - \tilde{A}_{k_{s,t}}^{s,t}\|_F &\leq \|A|_{s\times t} - \tilde{A}_{k_{s,t}-1}^{s,t}\|_F \\ &\approx \|\tilde{A}_{k_{s,t}}^{s,t} - \tilde{A}_{k_{s,t}-1}^{s,t}\|_F \\ &= \|v^{k_{s,t}}(w^{k_{s,t}})^T\|_2 \\ &= \|v^{k_{s,t}}\|_2 \|w^{k_{s,t}}\|_2. \end{aligned} \tag{23}$$

We also expect that normalization Eq. (19) reduces the condition number of the coefficient matrix. Multiplying $\|\phi\|_2$ to the numerator and the denominator of the left hand side of inequality Eq. (11) and using the relation $\tilde{a}\phi = a(\phi + \delta\phi)$ gives the following evaluations with condition number $\kappa(a)$,

$$\frac{\|\delta\phi\|_2}{\|\phi\|_2} \leq \varepsilon\|a^{-1}\|_2\|a\|_2 = \varepsilon\kappa(a), \tag{24}$$

$$\frac{\|a\delta\phi\|_2}{\|a\phi\|_2} \leq \varepsilon\kappa(a)^2. \tag{25}$$

These equations imply a relation between the condition number and the influence of the matrix approximation on vector ϕ. If we require ϕ or $a\phi$ to be within a certain degree of accuracy, the larger the condition number is, the smaller the tolerance ε we need to set.

Summarizing the discussion above, we propose a method for

calculating Eq. (2) with the approximated coefficient matrix described by Eq. (8). If we can easily obtain information about the sizes of the basis supports and determine the tolerance for the 2-norm error of matrix approximation, the approximated coefficient matrices $\tilde{a}^{s,t}_{k_{s,t}}$ are computed using ACA with the stopping criterion Eq. (16). If this is not the case, the matrix approximation is performed according to the following steps.

(A) Normalize Eq. (2) using Eqs. (18) and (19).

(B) Perform correctly clustering, e.g., according to the manner given in Ref. [3].

(C) Compute the approximated sub-matrices $\tilde{A}^{s,t}_{k_{s,t}}$ using ACA with the stopping criterion Eq. (22).

4. Numerical Experiments

We applied the proposed method and conventional H-matrices to the linear equation system derived from the quasi-dynamic equilibrium equation (QDEE) [16]. ACA+ is used as the low-rank approximation in both methods. The normalized Eq. (18) and stopping criterion Eq. (22) are adopted in our proposed method. Then, the constant C_{ov} is given by 1, since we use piecewise constant functions as base functions. For conventional H-matrices, we use the original Eq. (2) and the stopping criterion $\|v^{k_{s,t}}\|_2\|w^{k_{s,t}}\|_2 \leq \varepsilon\|\tilde{a}|_{s\times t}\|_F$, where the heuristic Eq. (23) and the replacement of $\|a|_{s\times t}\|$ to $\|\tilde{a}|_{s\times t}\|$ are used for the criterion Eq. (12). In the earthquake cycle simulation, the QDEE is utilized to calculate the stress shear from the physical quantities such as the slip, slip rate, and plate velocity of the continent [11]. The QDEE can be written in the form of Eq. (1) where the main term is an integral operator with the kernel function in the form of Eq. (5). The largest order of singularity of the kernel function is proportional to $|x-y|^{-3}$. As reported in Ref. [11], the computational time and memory usage increase rapidly as matrix size n increases when conventional H-matrices are applied to the QDEE. These increases are due to increasing the rank of the far-off-diagonal sub-matrix. The aim of the numerical experiments is to confirm that our proposed method can prevent the increase in rank.

As a preliminary step, we measure the relative error ε_H for the approximated matrix when setting tolerance ε_{ACA+} for the stopping criterion of ACA+. Then, $\varepsilon_H = \frac{\|a-\tilde{a}\|_F}{\|a\|_F}$ for conventional H-matrices, while $\varepsilon_H = \frac{\|A-\tilde{A}\|_F}{\|A\|_F}$ for our proposed method. The results with $\varepsilon_{ACA+} = 10^{-3}, 10^{-4}, 10^{-5}$ and matrix size $n = 11{,}520$ and 2,880 are given in **Table 1**. All the calculations in Table 1 are carried out by using serial computing on Fujitsu FX10 at The University of Tokyo, which consists of SPARC64$^{\text{TM}}$ IXfx and 32 GB memory in a node. From the point of view of controlling the accuracy of matrix approximation, it is optimal that $\varepsilon_H = \varepsilon_{ACA+}$. As the second best option, ε_H should be smaller than ε_{ACA+}. In both methods, the measured errors are much smaller than the given tolerances. Moreover, ε_H by conventional H-matrices decrease in proportion to ε_{ACA+}, while ε_H by our proposed method does not do in proportion. There is room for the improvement on controlling the accuracy. The calculated error ε_H by our proposed method is closer to ε_{ACA+}(the setting value) than that by conventional H-matrices. The difference between them is about an order of 10 in all cases.

Table 1 Relative errors by our proposed method and conventional H-matrices.

ε_{ACA+}	n=11,520		n=2,880	
	Conventional H-matrices $\varepsilon_H = \frac{\|a-\tilde{a}\|_F}{\|a\|_F}$	Our proposed method $\varepsilon_H = \frac{\|A-\tilde{A}\|_F}{\|A\|_F}$	Conventional H-matrices $\varepsilon_H = \frac{\|a-\tilde{a}\|_F}{\|a\|_F}$	Our proposed method $\varepsilon_H = \frac{\|A-\tilde{A}\|_F}{\|A\|_F}$
1.0E-3	6.2779E-7	8.0841E-06	5.2231E-07	3.7407E-06
1.0E-4	6.0116E-8	2.1486E-07	7.1935E-08	7.1034E-07
1.0E-5	6.1812E-9	2.5731E-07	8.3662E-09	4.4681E-07

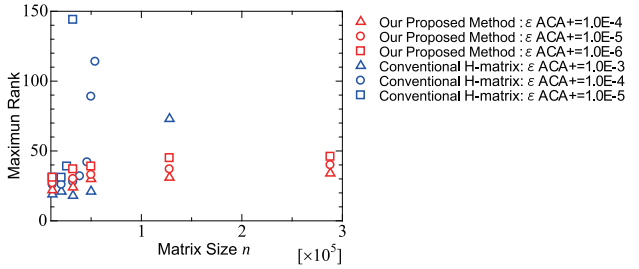


Fig. 2 Maximum rank $k_{\max}$ of all the ranks of approximated sub-matrices as a function of matrix size n. Although $k_{\max}$ increases rapidly as the matrix size increases when using conventional H-matrices, such an increase is not observed when using our proposed method.

We now investigate the behavior of the maximum rank $k_{\max}$ when varying matrix size n. Here, maximum rank $k_{\max}$ means the maximum of all the ranks of sub-matrices approximated by ACA+. In **Fig. 2**, the calculated $k_{\max}$ values by both methods are plotted as a function of matrix size n. Considering the above preliminary step, we suppose the ε_H by our proposed method is 10 times greater than that by conventional H-matrices in our examples. Therefore, we set $\varepsilon_{ACA+} = 10^{-3}, 10^{-4}, 10^{-5}$ for conventional H-matrices and $\varepsilon_{ACA+} = 10^{-4}, 10^{-5}, 10^{-6}$ for our proposed method. For all cases using conventional H-matrices, we can see a rapid increase in the maximum rank $k_{\max}$. Accordingly, the computational time and memory usage also increase. Even if we set a relatively large tolerance $\varepsilon_{ACA+} = 10^{-3}$ in the conventional method, a similar increase in $k_{\max}$ is observed when the matrix size n is greater than about 10,000. Furthermore, the rate of increase accelerates as ε_{ACA+} becomes smaller. With $n = 288{,}000$, ACA+ in conventional H-matrices fails to create approximated sub-matrices within rank $k = 150$. For a much larger n, ACA+ needs a very large rank k to satisfy the stopping criterion. This means that enormous memory usage and unrealistic computational time are required to perform the simulation using the conventional method. On the other hand, when using our proposed method, the rapid increase in $k_{\max}$ is not observed in any case with matrix size n less than 460,000. The rate of increase is very slow compared with the cases using conventional H-matrices. The increase in $k_{\max}$ is only about 6 when decreasing ε_{ACA+} to one tenth, independently of the matrix size. Even with $n = 288{,}000$, the ACA+ in our proposed method creates approximated sub-matrices within rank $k = 50$. Moreover, we can execute the simulation maintaining the benefits of H-matrices, but decreasing both the memory usage and the computational time.

Our main issue in this paper is to propose a remedy for the problem such that the conventional H-matrices with ACA fail to construct approximation by a rapid increase in $k_{\max}$. It is also an

Table 2 Memory usage and computational time when constructing an H-matrix by using our proposed method and conventional H-matrices.

matrix size	Conventional H-matrices			Our proposed method		
n	$k_{\max}$	memory (MB)	time (s)	$k_{\max}$	memory (MB)	time (s)
32,000	18	309	373	20	318	428
128,000	73	1,630	1,966	25	1,753	2,286
288,000	more than 1,000	—	more than 72,000	29	4,636	6,019

important issue that memory usage and the computational time of our proposed method are compared with ones of conventional H-matrices. The computational effort in the use of H-matrices depends on the number of entries $N_{\tilde{a}}$. As known by the expression Eq. (10) for $N_{\tilde{a}}$, the higher number of maximum rank $k_{\max}$ does not always result in the higher computational cost. If the size of sub-matrix with $k_{\max}$ is small, the influence on the total computational effort would not be significant. However, as mentioned in Section 1, the sub-matrix with $k_{\max}$ often appear at far-off-diagonal part when using the conventional H-matrices. The sub-matrix at far-off-diagonal part is ordinarily the largest sub-matrix, therefore the higher number of $k_{\max}$ can bring the big impact to the total computational effort. In addition to the investigation for $k_{\max}$ above, we examine the memory usage and the execution time when constructing an H-matrix by using our proposed and conventional H-matrices methods. We vary the matrix size by letting n = 32,000, 128,000, 288,000, and use $\varepsilon_{ACA+} = 10^{-3}$ in both approximation methods. The results are shown in **Table 2**. All the calculations in Table 2 are carried out by using serial computing on a computational node of CRAY XC30 at Kyoto University, which is equipped with Xeon E5™ and 64 GB memory. In the cases that conventional H-matrices work relatively well (n = 32,000, 128,000), the memory usage and the execution time of both methods are in the same order. However, in the case of n = 288,000, conventional H-matrices cannot construct an approximated matrix within 72,000 seconds, while our proposed method complete it in 6,019 seconds. Then, ranks of some sub-matrices at far-diagonal part exceed 1,000 when using conventional H-matrices. Such a rapid increase in ranks brings great influence to the memory usage and the execution time.

Finally we confirm whether the earthquake cycle simulation is suitably performed when the matrix in the QDEE is approximated by both our proposed method and conventional H-matrices. The slip velocity histories are computed using the above two approximation methods and the non-approximated original dense matrix. Matrix size n = 32,000 and tolerance $\varepsilon_{ACA+} = 10^{-4}$ are used in both approximation methods. The results are plotted as a function of time in **Fig. 3**. Each sharp peak in the slip velocity denotes the occurrence of an earthquake. We can see that the lines almost overlap and that all instances of earthquakes are in agreement.

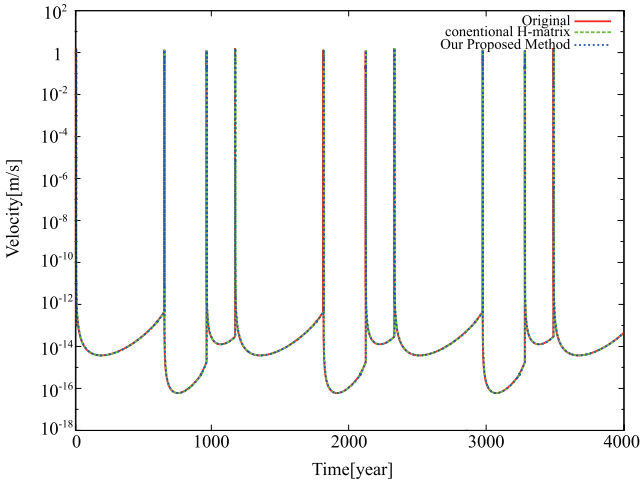


Fig. 3 Simulated slip velocity histories using our proposed method, conventional H-matrices, and the non-approximated original dense matrix with n = 32,000 and $\varepsilon_{ACA+} = 10^{-4}$. We plot each slip velocity history in the stable cyclic periods obtained after a few initial irregular cycles, and adjust the start time in the plot when a large instability occurs in each case. The lines almost overlap in all cases.

5. Conclusion

In this paper, we presented a method for calculating vast linear equation systems derived from integral equations whose kernel functions have singularities, especially high-order ones. The BEM is used in this case for discretizing the integral equation. The proposed method is based on H-matrices with ACA, which is known as one of the fastest techniques for such problems. In our proposed method, we introduced a kind of normalization for the linear equation system and a new stopping criterion for ACA used as the low-rank approximation of sub-matrices induced by H-matrices. These modifications were carried out to avoid the issue encountered by conventional H-matrices, whereby the maximum rank over all the ranks of approximated sub-matrices increases rapidly as the matrix size increases. Suppressing the rank of sub-matrices leads to reduced computational time and memory usage. We anticipate that our proposed method prevents the same issue caused by the non-uniformity of boundary element sizes. Although we began the formulation from an integral equation in this paper, our proposed method could be applied, as is, to the n-body problem.

As an example, our proposed method was applied to the linear equation system derived from a QDEE in an earthquake cycle simulation. It was shown that the earthquake cycle simulation could be suitably performed if the coefficient matrix was approximated by either our proposed method or conventional H-matrices. Furthermore, we confirmed that sub-matrices induced by H-matrices with our proposed method can be approximated, while at the same time suppressing the maximum rank in all cases where the matrix size n is less than 460,000. Our proposed method enables us to compute the huge data in an earthquake cycle simulation.

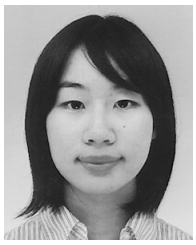
References

[1] Hackbusch, W.: A sparse Matrix Arithmetic Based on H-matrices. I. Introduction to H-matrices, *Computing*, Vol.62, No.2, pp.89–108 (1999).
[2] Hackbusch, W. and Khoromskij, B.N.: A Sparse H-matrix Arithmetic. II. Application to Multi-dimensional Problems, *Computing*, Vol.64, No.1, pp.21–47 (2000).
[3] Börm, S., Grasedyck, L. and Hackbusch, W.: Hierarchical Matrices, Lecture Note, Max-Planck-Institut fur Mathematik (2006).
[4] Greengard, L. and Rokhlin, V.: A New Version of the Fast Multipole Method for the Laplace Equation in Three Dimensions, *Acta Numerica*, Vol.6, pp.229–269 (1997).
[5] Cheng, H., Greengard, L. and Rokhlin, V.: A Fast Adaptive Multipole Algorithm in Three Dimensions, *J. Comput. Phys.*, Vol.155, pp.468–

498 (1999).
[6] Nishimura, N.: Fast Multipole Accelerated Boundary Integral Equation Methods, *Appl. Mech.*, Vol.55, No.4, pp.299–324 (2002).
[7] Barnes, J. and Hut, P.: A Hierarchical O(N log N) Force-calculation Algorithm, *Nature*, Vol.324, No.4, pp.446–449 (1986).
[8] Goreinov, S.A., Tyrtyshnikov, E.E. and Zamarashkin, N.L.: A Theory of Pseudoskeleton approximations, *Lin. Alg. Appl.*, Vol.261, pp.1–22 (1997).
[9] Kurtz, S., Rain, O. and Rjasanow, S.: The Adaptive Cross-Approximation Technique for the 3-D Boundary-Element Method, *IEEE Trans. Magn.*, Vol.38, No.2, pp.421–424 (2002).
[10] Bebendorf, M. and Rjasanow, S.: Adaptive Low-Rank Approximation of Collocation Matrices, *Comput-ing*, Vol.70, No.1, pp.1–24 (2003).
[11] Ohtani, M., Hirahara, K., Takahashi, Y., Hori, T., Hyodo, M., Nakashima, H. and Iwashita, T.: Fast Computation of Quasi-Dynamic Earthquake Cycle Simulation with Hierarchical Matrices, *Procedia Computer Science*, Vol.4 (2011).
[12] Ostrowski, J., Andjelic, Z. and Bebendorf, M.: Fast BEM-Solution of Laplace Problems with H-Matrices and ACA, *IEEE Trans. Magn.*, Vol.42, No.4, pp.627–630 (2006).
[13] Popović, N. and Praetorius, D.: H-Matrix Techniques for Stray-field Computations in Computational Micromagnetics, Large-Scale Scientific Computing Lecture Notes, *Computer Science*, Vol.3743, pp.102–110 (2006).
[14] Bradley, A.M.: H-matrix and Block Error Tolerances, arXiv1110.2807v1[math.NA] (2011).
[15] Golub, G.H. and Van Loan, C.F.: *Matrix Computations — Third Edition*, Baltimore, The Johns Hopkins University Press (1996).
[16] Rice, J.R.: Spatio-temporal Complexity of Slip on a Fault, *J. Geophys. Res.*, Vol.98, No.B6, pp.9885–9907 (1993).

Akihiro Ida was born in Japan in 1971. He received his B.Math and M.E. degrees from Nagoya University in 1994 and 1996, respectively. In 2008, Chuo University awarded him a Ph.D. degree in mathematics. He researched and developed linear solvers at VINAS Co., Ltd in 2000–2012. He currently engages in a Japanese national project of JST CREST as an assistant professor in ACCMS, Kyoto University. His research interests include discretization methods for integro-differential equations, magnetohydrodynamics and high performance computing.

Takeshi Iwashita was born in 1971. He received a B.E., an M.E., and a Ph.D. from Kyoto University in 1992, 1995, and 1998, respectively. In 1998–1999, he worked as a post-doctoral fellow of the JSPS project in the Graduate School of Engineering, Kyoto University. He moved to the Data Processing Center of the same university in 2000. In 2003–2014, he worked as an associate professor in the Academic Center for Computing and Media Studies, Kyoto University. He currently works as a professor in the Information Initiative Center, Hokkaido University. His research interests include high performance computing, linear iterative solver, and electromagnetic field analysis. He is a member of IEEE, SIAM, IPSJ, IEEJ, JSIAM, JSCES, and JSAEM.



Makiko Ohtani was born in Japan in 1987. She received her B.S. and M.S. degrees from Kyoto University in 2010 and 2012, respectively. Since 2010, she has been attending the graduate course at Kyoto University. Her research interests are in seismology, and techtonophysics.

Kazuro Hirahara was born in 1952. He received a B.S., and M.S. and Ph.D. from Kyoto University, in 1975, 1977, and 1981, respectively. In 1981 and 1982, he worked as a post-doctoral fellow of JPS project in Graduate School of Science, Kyoto University, and as a part-time lecture in Faculty of Science, Kyoto Sangyo University, respectively. Then, he worked an assistant professor in 1983–1990, and as an associate professor in 1990–1996, in Disaster Prevention Research Institute, Kyoto University. He moved to Graduate School of Science, Nagoya University, as a professor in 1996. He worked as a professor there in 1996–2001, and in Graduate School of Environmental Studies in 2001–2005. Then, in 2005, he moved to Graduate School of Science, Kyoto University as a professor, and currently works there. His research field is Seismology in Solid Earth Science, and his special interest includes simulations of earthquake cycles and of seismic waves traveling in the Earth for investigating the generation mechanism of earthquakes and the Earth's deep interior. He is a member of SSJ, GSJ, JGU, SEGJ, JSAFS and AGU.